

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2002, 20:17:55 ; Search time 3156 Seconds
(without alignments)
3061.512 Million cell updates/sec

Title: US-09-848-852A-3
Perfect score: 1772
Sequence: 1 MYSAQRFWGTIWARRGHAHLA.....DMPKLRRIYKELCHCKLTV 332

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=framet_p2n.model -DEV=xlp
-Q=/cgn2.1/USFTO_spool/US09848852/runat_16122002_132040_19132/app_query.fasta_1.519
-DB=GenEmbl -Qfmt=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09848852_QCGN_1.1.3637 @runat_16122002_132040_19132 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1751	98.8	1782	9	HSM800198	AL050283 Homo sapi
2	1751	98.8	2206	6	AX081143	AX081143 Sequence
3	1751	98.8	2224	9	AY008763	AY008763 Homo sapi
4	1751	98.8	2258	9	AF199459	AF199459 Homo sapi
5	1751	98.8	2269	9	HSM805332	AL834294 Homo sapi
6	1696	95.7	1707	10	AF194031	AF194031 Mus muscu
7	1696	95.7	2174	10	AY008764	AY008764 Mus muscu
8	969.5	54.7	139405	2	AC126239	AC126239 Felis cat
9	880	49.7	161428	2	AC126925	AC126925 Canis fam
10	849.5	47.9	153553	2	AC126921	AC126921 Bos tauru
11	820	46.3	1578	9	AX000923	AX000923 Homo sapi
12	794	44.8	257644	2	AC099294	AC099294 Rattus no
13	791	44.6	2007	9	AB060892	AB060892 Macaca fa
14	789	44.5	1830	9	BC008589	BC008589 Homo sapi
15	789	44.5	2793	9	BC030705	BC030705 Homo sapi
16	788.5	44.5	2673	9	AB074445	AB074445 Macaca fa
17	781.5	44.1	164504	2	AC129071	AC129071 Pan trogl
18	781.5	44.1	177703	2	AC016876	AC016876 Homo sapi
19	781.5	44.1	215795	2	AC127470	AC127470 Pan trogl
20	768	43.3	173588	2	AC127468	AC127468 Papio cyn
21	746	42.1	138792	2	AC119115	AC119115 Rattus no
22	726.5	41.0	76754	2	AC097861	AC097861 Rattus no
23	719	40.6	177555	2	AC130192	AC130192 Sus scrof
24	711.5	40.2	203083	2	AC069459	AC069459 Mus muscu
25	711.5	40.2	234182	10	AL603707	AL603707 Mouse DNA
26	691	39.0	203281	2	AC126237	AC126237 Canis fam
27	674	38.0	184026	2	AC098923	AC098923 Rattus no
28	661.5	37.3	503	6	AX209870	AX209870 Sequence
29	658	37.1	251835	2	AC099436	AC099436 Rattus no
30	566.5	32.0	333300	2	AC125091	AC125091 Mus muscu
31	519	29.3	209237	2	AL806532	AL806532 Mus muscu
32	474.5	26.8	5318	6	AX069229	AX069229 Sequence
33	474.5	26.8	6732	9	AF175325	AF175325 Homo sapi
34	473.5	26.7	164504	2	AC129071	AC129071 Pan trogl
35	456	25.7	1932	9	AF149770	AF149770 Homo sapi
36	456	25.7	2512	6	AX081135	AX081135 Sequence
37	453.5	25.6	2903	10	BC023129	BC023129 Mus muscu
38	396.5	22.4	150136	10	AC122027	AC122027 Mus muscu
39	396.5	22.4	267305	2	AC125216	AC125216 Mus muscu
40	395	22.3	182718	2	AC103148	AC103148 Rattus no
41	393	22.2	2484	9	AK074357	AK074357 Homo sapi
42	393	22.2	2497	9	AF151697	AF151697 Homo sapi
43	393	22.2	2795	6	AX081141	AX081141 Sequence
44	393	22.2	3031	9	HSM805453	AL834380 Homo sapi
45	393	22.2	3133	6	BD012321	BD012321 Genes rel

ALIGNMENTS

RESULT 1

HSM800198 1782 bp mRNA linear PRI 10-MAR-2001
 LOCUS Homo sapiens mRNA; cDNA DKF2p586K0919 (from clone DKF2p586K0919);
 DEFINITION complete cds.
 ACCESSION AL050283
 VERSION AL050283.1 GI:4886466
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1782)
 Wiemann,S., Well,B., Wellenreuther,R., Gassenhuber,J., Glassl,S.,
 Ansoerg,W., Boecker,M., Bloeker,H., Bauersachs,S., Blum,H.,
 Lauber,J., Duesterhoeft,A., Beyer,A., Koehler,K., Strack,N.,
 Mewes,H.W., Ottenwaelder,B., Obermaier,B., Tampe,J., Heubner,D.,
 Wambutt,R., Korn,B., Klein,M. and Poustka,A.
 Toward a Catalog of Human Genes and Proteins: Sequencing and
 Analysis of 500 Novel Complete Protein Coding Human cDNAs
 Genome Res. 11 (3), 422-435 (2001)
 11230166
 2 (bases 1 to 1782)
 Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
 Direct Submission
 Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152
 Martinsried, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 This clone (DKF2p586K0919) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at <http://www.mips.biochem.mpg.de/proj/cDNA/>.
 FEATURES
 Location/Qualifiers
 1..1782
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKF2p586K0919"
 /tissue_type="uterus"
 /clone_lib="586 (synonym: hutel). Vector pSport1; host
 DH10B; sites NotI + SalI/MluI"
 /dev_stage="adult"
 204..1319
 /gene="DKF2p586K0919"
 204..1319
 /gene="DKF2p586K0919"
 /note="similarity to S.pombe SPBC19G7.09"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAB43384.1"
 /db_xref="GI:4886467"
 /db_xref="SPTREMBL:Q9Y3W9"
 /translation="MAEDGVRSPPVSGPPMEGLRWTPKSLPDPDGLLSCTLPN
 GFQSGGEGRSAPPDASILISNVCSIGDHVAQELFQSDGLMAEAEPRGKAGQ
 HSPLEERHVTQVSTLDFLOTGLSLPLSTDEVVEKLEDFQOEFSPSRKGLVLQI
 IOSVIRPGNVAWGFGRVAYKRHLVTMDDLGLYCONWLNQVNMVYGDLYMDTVPEK
 VHFNSFFYDKLRKGYGVKRWTKNVDIFNKKLLLPILHLEVHWSLISVDVRRRTIT
 YFDSQRTLNLRKPHIAKYLQAEAVKKRDLDFHOGWKGFKMNVARQNDSDCGAFVL
 QYCKHLALSQPSFTQQDMPKLRQIYKELCHCKLTV"
 1741..1746
 polyA_signal
 polyA_site
 BASE_COUNT 419 a 466 c 475 g 422 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 8e-165 Length: 1782
 Score: 1751.00 Matches: 331
 Percent Similarity: 99.40% Conservative: 0
 Best Local Similarity: 99.40% Mismatches: 1
 Query Match: 98.81% Indels: 1
 DB: 9 Gaps: 0

US-09-848-852a-3 (1-332) x HSM800198 (1-1782)
 QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20
 Db 320 ATGTACTCTGCCCAACGGTTTTGGGGGACAATCTGGCCAGAAAGGGGAGCGACGCTGGC 379
 QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
 Db 380 ACCCCCTGATGCGCAGCATCCTCATCAGCAATGTGTGCACATCGGGGACCATGTGGCCCA 439
 QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGlu 60
 Db 440 GGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAAAGGAGGAGAGCGCTGGGAGAA 499
 QY 60 salAGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
 Db 500 AGCCGCCAGCAGACGCCCTCGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTGA 559
 QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGlu 100
 Db 560 CGAATCTCTTAACAGTATGGCAGCCTCATACCCCTCAGCACTGATGAGGTAGTAGAA 619
 QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGl 120
 Db 620 GCTGGAGGACATTTTCCAGCAGGAGTTTTCCACCCCTTCCAGGAAGGGCCTGGTGTGCA 679
 QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
 Db 680 GGTGATCCAGCTTTACCAGCGATGCCAGCAATGCCATGGTGAGGGGCTTCCGAGTGCG 739
 QY 140 aTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160
 Db 740 TTATAGCGGCGAGTGTGCTGACCATGTGATGACTGGGACCTGTATGGACAGAACTGGCT 799
 QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
 Db 800 CAATGACCAGGTGATCAACATGTATGGAGACCTGTGTGACACACAGTCCCTGAAAGGT 859
 QY 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
 Db 860 GCATTTCTTCAATAGTGTCTTCTATGATAAATCCGTACCAAGGGTTATGATGGGGTGAA 919
 QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220
 Db 920 AAGTGGACCAAAACGTTGGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACCT 979
 QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
 Db 980 GGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTAGGGCGACGACCATCACCTATTTTGA 1039
 QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260
 Db 1040 CTCGCGAGGTACCCAAAACCGCGCTGCGCTAAGCATATTCCAAGATATCTCAGAGCAGA 1099
 QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280
 Db 1100 GCGGTGAAGAAGACCGACTGGATTTCCACCAGGCTGGAAGGTACTTCAAAATGAA 1159
 QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300
 Db 1160 TGTGCCAGGCAGAAATAATGACAGTGACTGTGCTGTTGTTGTGTGCTACTGCAAGCA 1219
 QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgGl 320
 Db 1220 TCTGGCCCTGTCTCAGCCATTTCAGCTTACCCAGCAGGACATGCCCAAACCTTCGTGCGCA 1279
 QY 320 nIleTyrLysGluLeuCysHisCysLysLysLeuThrVal 332
 Db 1280 GATCTACAAGGAGCTGTGTCACCTGCAAACTCACTGTG 1316
 RESULT 2
 AX081143

LOCUS AX081143 2206 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 9 from Patent WO0109292.
ACCESSION AX081143
VERSION AX081143.1 GI:13170039
KEYWORDS human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 2206)
Gong, L. and Yeh, E.T.
JOURNAL Composition and methods relating to senp1 - a sentrin-specific
protease
Patent: WO 0109292-A 9 08-FEB-2001;
Board of Regents, The University of Texas System (US)
FEATURES
source
1..2206
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 491 a 612 c 605 g 498 t
ORIGIN
Alignment Scores:
Pred. No.: 1.08e-164 Length: 2206
Score: 1751.00 Matches: 331
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 1
Query Match: 98.81% Indels: 1
DB: 6 Gaps: 0
US-09-848-852A-3 (1-332) * AX081143 (1-2206)
Qy 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHis-LeuAl 20
Db 759 ATGTACTGTCCCAACGGTTTGGGGACAATCTGGCGAAGAGGGAGCGAGCTTGGC 818
Qy 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1 40
Db 819 ACCCCTCATGCCAGCATCTCATCAGCAATGTGTGCAGCATCGGGACCATGTGGCCCA 878
Qy 40 pLutePheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60
Db 879 GGAGCTTTTTCAGGCTCAGATTTGGCGATGCGACAGAGGAGGAGCGCTTGGGGAGA 938
Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 939 ACCCGCCAGACACACCCCTCGGAGAGGAGCATGTGACCTGCGTACAGACATCTTGA 998
Qy 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValAlaGluLy 100
Db 999 CGAATCTCTCAAGGTATGCGAGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGAA 1058
Qy 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuG1 120
Db 1059 GCTGGAGACATTTTCCACAGAGGAGTTTCCACCCCTCCAGGAAGGCGCTGGTGTGCA 1118
Qy 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 1119 GCTGATCCAGTCTTACCAGCGATGCCAGGCAATGCCATGTGTGAGGGCTTCCGAGTGGC 1178
Qy 140 aTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpIle 160
Db 1179 TTATAAGCGGCACGCTGCTGACCATGGATGACTTGGGGACCTGTATGGACAGAACTGGCT 1238
Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLyVa 180
Db 1239 CAATCACCAGCTGATGAACATGTATGGAGACCTGCTCATGCACACAGTCCCTGAAAGGT 1298
Qy 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
Db 1299 GCATTTCTTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTAA 1358
Qy 200 sArgTrpThrLysAsnValAspIlePheAsnValLysGluLeuLeuLeuIlePheHisLe 220

|||||
Db 1359 AAGGTGGACAAAACGTTGGACATCTTCAATAAGGAGCTACTGTCTAAATCCCATCCACT 1418
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
Db 1419 GGAGTGCATTGGTCCCTCATCTCTGTGTGATGTGAGGCGACGACCATCACTATTTGA 1478
Qy 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaG1 260
Db 1479 CTCGAGCGTACCTTAACCCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCGAGA 1538
Qy 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280
Db 1539 GCGGCTAAAGAAAGACCGACTGGATTTCACACGCGCTGGAAAGGCTTACTTCAAAATGAA 1598
Qy 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300
Db 1599 TGTGGCCAGGAGAAATAATGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1658
Qy 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgG1 320
Db 1659 TCTGGCCCTGTCTCAGCCATTTCAGCTTCCACGAGGAGACATGCCCAACTTCGTGCGCA 1718
Qy 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
Db 1719 GATCTACAAGGAGCTGTGTCTACTGCAAACTCACTCTGTG 1755
RESULT 3
AY008763
LOCUS
DEFINITION Homo sapiens sentrin/SUMO-specific protease (SEN3), mRNA, complete
cds.
ACCESSION AY008763
VERSION AY008763.1 GI:11245810
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 2224)
AUTHORS Yeh, E.T., Gong, L. and Kamitani, T.
TITLE Ubiquitin-like proteins: new wines in new bottles
JOURNAL Gene 248 (1-2), 1-14 (2000)
MEDLINE 20267842
PUBMED 10806345
REFERENCE 2 (bases 1 to 2224)
AUTHORS Gong, L. and Yeh, E.T.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2000) Institute of Molecular Medicine, University
of Texas-Houston, 2121 W. Holcombe Blvd. 928, Houston, TX 77030,
USA
FEATURES
source
1..2224
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p13"
1..2224
/gene="SEN3"
52..1776
/gene="SEN3"
/codon_start=1
/product="sentrin/SUMO-specific protease"
/protein_id="AAG33252.1"
/db_xref="GI:11245811"
/translation="MKETIQGTSGWGPPEPPGIPPAYSSPRDGLRWPPPPKPLKLS
GGGPDGPGSGTTPTRRLPAPRDFASASEEEEEDEDEDEEVEEVAARLPPRW
SGLGTSORPRALRPSHRKTCSPORERRAMFOMLLYSKSTSLTFHKLWGRHRRRG
LAHPKHLSPQGGATQVPSPCRFSPGPPPRGLGLGALMAEDGVGRGSPVPSG
PPMEEDGLRWTPKSPDLDSGLLSCTLPNGFGGSGGEGERSLAPPDASILLINCS1
GDHVAELFGQSDILGMAEAEAPGKAGHSPLREHVTVCQSLDLEFLQYSLIPL
STDVVEKLEDIFQOEFSTPSRKGLVLQIQSVQRMFGNAMVRGFRVAYKRHLTMD
LGTLYGQNLNDQVMNMYGDLVMDTVPEKVHFFNSFFYDKLRTKGYDGVKRWTKNVDI

FNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRCPKHIKYLQAEAVKKDR
LDFHQGWKGYPKMNVARQNDSDCGAFVLQYCKHLALSQPFSTQDDMPKLRRIYKE
LCHCKLTV"

BASE COUNT 502 a 615 c 612 g 495 t
ORIGIN

Alignment Scores:
Pred. No.: 1,09e-164 Length: 2224
Score: 1751.00 Matches: 331
Percent Similarity: 99.40% Conservative: 0
Best local Similarity: 99.40% Mismatches: 1
Query Match: 98.81% Indels: 1
DB: 9 Gaps: 0

US-09-848-852A-3 (1-332) x AY008763 (1-2224)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20
DB 777 ATGTACTCTGCCAACGGTTTGGGGACAATCTGGCCAGAGGGGAGCGCAGCTTGGC 836
QY 20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
DB 837 ACCCTGTATGCGAGCATCTCATAGCAATGTGTGAGCATCGGGGACCATGTGGCCCA 896
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGlu 60
DB 897 GGAGCTTTTTCAGGCTCAGATTTGGCATGGCAGAGAGGAGGAGGCTGGGAGAA 956
QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
DB 957 AGCCGCCACGACAGCCCTGGGAGAGGAGCATGTGACCTGGGTACAGAGCATCTTGA 1016
QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValGlu 100
DB 1017 CGAATTCCTCAACAGTATGGCAGCTCATACCCCTCAGACACTGATGAGTAGAGAA 1076
QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGl 120
DB 1077 GCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAGGCGCTGGTGTGCA 1136
QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
DB 1137 GCTGATCCAGTCTTACAGCGGATGCGAGCAATGCCATGGTGAGGGGCTTCGAGTGGC 1196
QY 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160
DB 1197 TTATAAGCGCAGCTGCTCACCATGATGACTTGGGGACCTTGTATGGACAGACTGGCT 1256
QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGlyLysVa 180
DB 1257 CAATGACCGAGTGATCAACATGATGAGACCTGGTCTATGGACACAGTCCCTGAAAGGT 1316
QY 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
DB 1317 GCATTTCTTCAATAGTTTCTTATATATAAATCCCTACCAAGGGTTATGATGGGTGAA 1376
QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuIleProIleHisLe 220
DB 1377 AAGTGGACCAAAAGCTGGACATCTTCAATAAGGAGCTACTGTAAATCCCATCCACCT 1436
QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
DB 1437 GGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTGAGCGCGACGACCATCACCTATTTC 1496
QY 240 pSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260
DB 1497 CTGCGACGCTACCTTAACCGCGCTTCCGCTAAGCATATTGCCAAGTATCTACAGCAGA 1556
QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280
DB 1557 GCGGTAAAGAAGACCGACTTGATTTCCACAGGCGCTGGAAGGGTACTTCAAAATGAA 1616
QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysH 300

DB 1617 TGTGGCCAGCGAATAATGACAGTGCAGTGTGTGCTTTTGTGTGCTACTGCAAGCA 1676
QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGl 320
DB 1677 TCTGGCCCTGTCTCAGCCATTGAGCTTCACCCAGCAGGACATGCCCAAACTTCGTGGCA 1736
QY 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
DB 1737 GATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTG 1773
RESULT 4
AF199459
LOCUS Homo sapiens SUMO-1 specific protease 3 (SSP3) mRNA, complete cds.
DEFINITION AF199459
ACCESSION AF199459
VERSION AF199459.1 GI:16550942
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2258)
AUTHORS Choi,S.J., Jeon,Y.J., Kim,K.I., Nishimori,S., Suzuki,T., Uchida,S.,
Shimbara,N., Tanaka,K. and Chung,C.H.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1999) Molecular Biology, Seoul National
University, Shillim-dong san 56-1, Seoul 151-742, Korea
FEATURES
source
1..2258
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene
1..2258
/gene="SSP3"
CDS
71..1795
/gene="SSP3"
/codon_start=1
/product="SUMO-1 specific protease 3"
/protein_id="AAL25652.1"
/db_xref="GI:16550943"
/translation="MKETIQGTGSGWPPGPGIPPAYSSPRERLRWPPPPKPRLS
GGGPGDPGSGTTVPARLPVPRFSFASAEEDDEEEDEEEDEEEAAWRLPPRW
SQLGTSQRPSRPTHRKTCORRRAMRFLYLSKSTSLTFHWKLGRRHRRRRG
LAHPKNHLSPOQGATPOVPPQCRFRSPPPRLGLGALMAEDVRGSPVPVSG
PMEDGLRWTPKSPLDPSGLISCTLPNGFGGSGPGRSLAPPDASILISNCSI
GDHVAQELFGSDGLGMAEAEAPGEEKAGQHSPLREHVTQVSLDEFLQYGSILPL
STDEVEKLEDFIQEFTSPSRKGLVLQISYQMPGNAMVRFVAYKRRHLTMD
LGTLYGQNLNDQVMNYGDLVMDTVPKVFHFFNSFFYDKLRTKGYDGVKRWTKNDI
FNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRCPKHIKYLQAEAVKKDR
LDFHQGWKGYPKMNVARQNDSDCGAFVLQYCKHLALSQPFSTQDDMPKLRRIYKE
LCHCKLTV"
BASE COUNT 519 a 626 c 618 g 495 t
ORIGIN
Alignment Scores:
Pred. No.: 1,12e-164 Length: 2258
Score: 1751.00 Matches: 331
Percent Similarity: 99.40% Conservative: 0
Best local Similarity: 99.40% Mismatches: 1
Query Match: 98.81% Indels: 1
DB: 9 Gaps: 0
US-09-848-852A-3 (1-332) x AF199459 (1-2258)
QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20
DB 796 ATGTACTCTGCCAACGGTTTGGGGACAATCTGGCCAGAGGGGAGCGCAGCTTGGC 855
QY 20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
DB 856 ACCCTGTATGCGAGCATCTCATAGCAATGTGTGACATCGGGGACCATGTGGCCCA 915
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGlu 60

Db	916	GGAGCTTTTTCAGGCTCAGATTGGGCATGGCAGAGAGCAGAGAGCCCTGGGGAA	975		
Qy	60	salaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs	80		
Db	976	AGCCGGCCAGCAGACAGCCCTTGCAGAGAGCATGTGACCTGCGTACAGACATCTTGA	1035		
Qy	80	pGluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValValGluLy	100		
Db	1036	CGAATTCCTTCAACAGTATGCAGCCTCATACCCCTCAGCACTGATGAGGTAGTAGAA	1095		
Qy	100	sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArglyGlyLeuValLeuG1	120		
Db	1096	GCTGGAGGACATTTTCCAGCAGAGAGTTTTCACCCCTTCCAGGAAGGCGCTGGTTCGA	1155		
Qy	120	nLeuIleGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArgValal	140		
Db	1156	GCTGATCCAGCTTACACAGCGATGCCAGGCAATGCCATGTGAGGGCTTCCGAGTGGC	1215		
Qy	140	aTyrlsArgHisValLeuThrMetAspLeuGlyThrLeuTyrglyGlnAsnTrpLe	160		
Db	1216	TTATAAGCGGCACGTGTCACCATGGATGACTTGGGACCTTGTATGGACAGAACTGGCT	1275		
Qy	160	uAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrValProGluLysVa	180		
Db	1276	CAATGACCAAGTGTGAACATGTATGGAGACCTGTCATGCACACAGTCCCTGAAAGGT	1335		
Qy	180	lHisPhePheAsnSerPhePheTyrglyAspLeuArgThrTyrglyTyrglyValLy	200		
Db	1336	GCATTTCTTCAATATGTTCTTATGATAAATCCGTAACCGTATGATGGGGTAA	1395		
Qy	200	sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuProIleHisLe	220		
Db	1396	AAGTGGACCAAAACGTGGACATCTCAATAAGAGCTACTGTATCCCATCCACCT	1455		
Qy	220	uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrglyPheAs	240		
Db	1456	GGAGGTGCATTGGTCCCTCATCTCTGTGTGATGTGAGGCGAGCCACCATCATTTTGA	1515		
Qy	240	pSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrglyLeuGlnAlaG1	260		
Db	1516	CTCCAGCGTACCCTAAACCCGCCCTGCCCTTAAGCATATCCCAAGTATCTACAGGCAGA	1575		
Qy	260	uAlaValLysLysAspArgLeuAspPheHisGlnGlyTyrglyTyrglyPheLysMetAs	280		
Db	1576	GGCGTAAAGAAAGACCGACCTGGATTTCCACAGCGCTGGAAGGTACTTCAAAATGAA	1635		
Qy	280	nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrglyLysH1	300		
Db	1636	TGTGGCCAGGAGAATAATGACAGTACTGTGTGTCTTTGTGTGAGTACTGCAAGCA	1695		
Qy	300	sLeuAlaIleuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgG1	320		
Db	1696	TCTGGCCCTGTCTCAGCATTTACCTGCTTACCCAGCAGACATGCCAAACTTCGTCGCA	1755		
Qy	320	nIleTyrlsGluLeuCysHisCysLysLeuThrVal	332		
Db	1756	GATCTACAGAGGTGTGTCTACTGCAAACTCATCTGTG	1792		
RESULT 5					
HSM805332					
LOCUS					
DEFINITION Homo sapiens mRNA; cDNA DKF2p762A152 (from clone DKF2p762A152).					
ACCESSION AL834294					
VERSION AL834294.1 GI:21739883					
KEYWORDS					
SOURCE human.					
ORGANISM Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; HomInidae; Homo.					
REFERENCE					
AUTHORS Koehrer K., Beyer A., Mewes H.W., Weil B. and Wiemann S.					
TITLE Direct Submission					
JOURNAL					
COMMENT Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY					
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer					
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;					
sequenced by BMFZ (Biomedical Research Center at the Charite,					
Berlin/Germany) within the cDNA sequencing consortium of the German					
Genome Project.					
This clone (DKF2p762A152) is available at the RZPD in Berlin.					
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059					
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further					
information about the clone and the sequencing project is available					
at http://mips.gsf.de/proj/cDNA/.					
FEATURES					
source					
1. 2269					
Location/Qualifiers					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/map="17"					
/clone="DKF2p762A152"					
/tissue_type="melanoma (MeWo cell line)"					
/clone_lib="762 (synonym: hm12). Vector pSport1; host					
DH10B; sites NotI + SalI"					
/dev_stage="adult"					
1. 2269					
/gene="DKF2p762A152"					
78. 1802					
/gene="DKF2p762A152"					
/note="partial similarity to sentrin/SUMO-specific					
protease (Homo sapiens)"					
/codon_start=1					
/product="hypothetical protein"					
/protein_id="CAD38967.1"					
/db_xref="GI:21739884"					
/translation="MKFTIOGTGSGWPPGPPGIPPAYSSPRRLRWPPPPKPRKLS					
GGGFGDPPGSGTTPVARRLPVPRSPDASASEEEEEDEDEEEVAARLPPPW					
SQGTSSRPSPRPTHRKTSQRRRAWRARFRLMLLSKSLTFHMLKMGHRRRRG					
LAHPKHLSPQGGATQVPSGCCFDPGPPPLRLGLLGLMAEDVGRGSPVPSG					
PMEEADGLRTPKSLDPSGLLSTLPNGFGSGSGEGRSLAPPDASLISNVCST					
GDHVAQLFQSGDLGMAEAEERGERAGHSPLREHVTCSQSLDFLQTYGSLIPL					
STDVYVKLEIDIFQOEPSTPSRKLVLQIOSYORMPGNAVRGFRVAYKRHLVTMDI					
LCVLYGONLWINDOMNMYGDLVMDTVEKVFHFNFSFYDKLRTKGYDGVKRWTKNVD					
FNKELLIIPIHLEWHWSLISVDVRRRTITYDSQRTLNRRCPKHIAYLQAEVKKDR					
LDPHQMGKGYFKMNAQNNDSDCGAFVLOYCKHLASQPSFTQDMPKLRIIYKE					
LCHCKLTIV"					
polyA_signal					
2228. 2233					
/gene="DKF2p762A152"					
polyA_site					
2252					
BASE COUNT					
ORIGIN					
522 a 628 c 621 g 498 t					
Alignment Scores:					
Pred. No.:					
Score:					
Percent Similarity:					
Best Local Similarity:					
Query Match:					
DB:					
US-09-848-852A-3 (1-332) x HSM805332 (1-2269)					
Qy					
1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHis-LeuAl					
Db					
803 ATGTACTCTGCCCAACGCTTTTGGGGACATCTGGGCCAGAGGGAGCCAGCTTGGC					
Qy					
20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1					
Db					
863 ACCCCCTGATGCCAGCATCTCATCAGCAATGTGTGCAGCATCGGGACCATGTGGCCA					
Qy					
40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGArgProGlyGluLy					
Db					
923 GGAGCTTTTTCAGGCTCAGATTGGGGCATGGCAGAGGAGGAGGAGCCCTGGGGAA					
Qy					
60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs					

```
Db 983 AGCGGCCACACAGCCCCCTGGCAGAGGAGCATGTGACCTGGCTACAGCAGCATCTTGA 1042
QY 80 pGluPheLeuGlnThrTyrglySerLeuileProLeuSerThrAspGluValValGluLy 100
Db 1043 CGAATTCCTTCAAAAGTATGGCAGCCTCATACCCCTCAGCAGCTGATGAGTAGAGAA 1102
QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuG1 120
Db 1103 GCTGGAGGACATTTCCAGCAGGAGTTTCCACCCCTCCAGGAGGCGTGGTGTGCA 1162
QY 120 nLeuileGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 1163 GCTGATCCAGCTCTACAGCGGATGCGAGCAATGCCATGGTGGAGGGCTCCGAGTGGC 1222
QY 140 aTyrgLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrgLysGlnAsnTrpLe 160
Db 1223 TTATAAGCGCAGCTGCTGACCATGATGACTTGGGAGCTTGTATGGACAGAACTGGCT 1282
QY 160 uAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrValProGluLysVa 180
Db 1283 CAATGACCAAGTATGAACATGTATGAGACCTGGTCATGGACACAGTCCTGAAAGGT 1342
QY 180 lHisPheAsnSerPhePheTyrglyAspLysLeuArgThrLysGlyTyrglyVally 200
Db 1343 GCATTTCTTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGTTATGATGGGTGAA 1402
QY 200 sArgTyrThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220
Db 1403 AAGTGGACCAAAACGTGGACATCTTCAATAGGAGCTACTGCTAATCCCATCCACCT 1462
QY 220 uGluValHisTrpSerLeuileSerValAspValArgArgThrIleThrTyrgPheAs 240
Db 1463 GGAGGTGCATTTGGTCCCTCATCTCTGTGATGGAGCGCAGCACCATCACCTATTTGA 1522
QY 240 pSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrgLysGlnAlaG1 260
Db 1523 CTCGCAGCTACCTTAACCGCGCTGCGCTAAGCATATTTGCCAAGTATCTACAGGCAGA 1582
QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrgPheLysMetAs 280
Db 1583 GCGCGTAAAGAAGACCGACTGATTTCCACAGGCGTGAAGGTTACTTCAAAATGAA 1642
QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrgCysLysHi 300
Db 1643 TGTGGCCAGCAGAAATATGACAGTACTGTGGTCTTTGTGTCAGTACTGCAAGCA 1702
QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgG1 320
Db 1703 TCTGGCCCTGTCTCAGCCATTGAGCTTCCAGCAGGAGCATGCCCAAACTTCGTGGCA 1762
QY 320 nIleTyrgLysGluLeuCysHisCysLysLeuThrVal 332
Db 1763 GATCTAAGGAGCTGTGCTGACTGCAAACTCACTGTG 1799

RESULT 6
AF194031 1707 bp mRNA linear ROD 01-NOV-2000
LOCUS
DEFINITION Mus musculus SMT3 isopeptidase 1 (Smt3ip) mRNA, complete cds.
ACCESSION AF194031
VERSION AF194031.1 GI:11066007
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1707)
Nishida,T., Tanaka,H. and Yasuda,H.
A novel mammalian Smt3-specific isopeptidase 1 (SMT3IP1) localized
in the nucleolus at interphase
Eur. J. Biochem. 267 (21), 6423-6427 (2000)
JOURNAL
MEDLINE 20485517
PUBMED 11029585
REFERENCE
2 (bases 1 to 1707)
```

```
AUTHORS Nishida,T., Tanaka,H. and Yasuda,H.
TITLE Direct Submission
JOURNAL Submitted (13-Oct-1999) School of Life Science, Tokyo University of
Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo
192-0392, Japan
FEATURES
Source
1..1707
/organism="Mus musculus"
/db_xref="taxon:10090"
gene
1..1707
/gene="Smt3ip"
CDS
1..1707
/gene="Smt3ip"
/codon_start=1
/product="SMT3 isopeptidase 1"
/protein_id="AAG28418.1"
/db_xref="GI:11066008"
/translation="MKETIQGTGSMGPEPPGPTTYSNPRERLRWLPKPKRLKSGG
FGPPDGGSTTVPTRRPSPFDASASEEEEEEEDEEVAARLPPRRNLGAS
QRSRALRPSHRKTCSSRRRARAFQMLLYSKSTLTFHWKLMGRHRRNLAPKED
HLSPOEGGATQVPSPCCFDPSRGLPPRLGLLALMAEDMGSPVPVSGPPMKEE
GLRWTPKSPDPDGLSLLCTLPNGFGGLSGPEGRSLAPPDASILLISNVCSIGDHVAQ
ELFSSDLGIAEADRTGEKAGQHSPLREEHVTCQSILDEFQTYIGSLIPSTDEVV
EKLEDIFQOEFSTPSRKSILVQLIQSYQRMFGFNAMRGFRVSKRHVLTMDLDTLYG
QNLNDQVMNMYGLVMDTVPEKVFHFNFSYDKLRTKGYDGVKRWTKNVDIFNKELL
LIPILHEVHWSLISVDVRRRTITVFDQSORTLNRRCPKHIAKYLOAEAVKKRLDFHOG
WKGYFKMNVARQNDSDCGAFVLQYCKHIALSOPSFSTTQQDMPKLRROIYRELCHCKL
TV"
BASE COUNT 406 a 477 c 461 g 363 t
ORIGIN
Alignment Scores:
Pred. No.: 2,28e-159 Length: 1707
Score: 1696.00 Matches: 320
Percent Similarity: 97.60% Conservative: 5
Best Local Similarity: 96.10% Mismatches: 7
Query Match: 95.71% Indels: 1
DB: 10 Gaps: 0
US-09-848-852A-3 (1-332) x AF194031 (1-1707)
QY 1 MetTyrgSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20
Db 708 ATGTACTCTACCAATGGTTTGGGGGACTGCTGCTGCTGAAGGGGAGCGAGCTGCG 767
QY 20 aProPheAlaSerIleLeuileSerAsnValCysSerIleGlyAspHisValAlaG1 40
Db 768 ACCCCCTGTATGCCAGCATACTCATCAGTAATGTGTGCAGCATTTGGAGACCAGCTCA 827
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60
Db 828 GGAATTTTTCAGAGCTCTGACTTGGGCATTCGCAAGAGGAGCATCGGACTGGGAGAA 887
QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 888 ACCTGGCCAGCATAGCCCTTCGGGAGGAGAACATGTGACCTGTGTGCAGAGTATCTAGA 947
QY 80 pGluPheLeuGlnThrTyrgLysLeuileProLeuSerThrAspGluValValGluLy 100
Db 948 TGAATTCCTTCAAACTTATGGCAGGCTCATCCCTCTCAGCAGCTGATGAAGTTGTAGAA 1007
QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuG1 120
Db 1008 GTTGGAGGACATTTTCCAGCAGGAGTTCTCTACACCTCAAGAGAGAGTCTGCTACTACA 1067
QY 120 nLeuileGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 1068 GCTGATCCAGTCTTATCAGCGGATGCCAGGCAACGCTATGTGTAAGGGGCTTCCGGGTATC 1127
QY 140 aTyrgLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrgLysGlnAsnTrpLe 160
Db 1128 CTATAAGGCAGACGCTGCTCACCATGATGACTTGGGTACCTTATATGGACAGAACTGGCT 1187
```

```

Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
Db 1188 CAATGACCAAGGTGATGAACATGATGAGATCTGCTGATGCACACAGTCCCTGAAAGGT 1247
Qy 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
Db 1248 GCATTCTTCAACAGCTCTTCTATGATGATAAATCCGTACCAAGGGTTATGATGGGGTAAA 1307
Qy 200 sArgTTrpThrLysAsnValAspIlePheAsnLysLeuLeuLeuLeuIleProIleHisLe 220
Db 1308 GAGGTGGACCAAAAATGTGGACATCTTCAATAAGGAATTAAGTCTTAATCCCATCATCT 1367
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
Db 1368 GGAGGTGCACCTGGTCCCTTATCTATCTAGTTGATGTAAGGCGAGTACCATCACCTATTTTGA 1427
Qy 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260
Db 1428 CTCCAGCGAACTCTAAATCCCGCTGCCCTTAAGCATATTGCCAAGTATCTACAGGCAGA 1487
Qy 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280
Db 1488 GGCAGTCAAAAAGACCCAGCTGGACTCCATCAGGGCTGGAAAGGTACTTCAAAATGAA 1547
Qy 280 nValalaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300
Db 1548 TGTGCCAGGCAGCAATAATGACAGTACTGTGGTGGCTTTGTATTACAGTACTGCAAGCA 1607
Qy 300 sLeuAlaLeuSerGlnPropheSerPheThrGlnGlnAspMetProLysLeuArgG1 320
Db 1608 CCTGGCCCTGTCTCAGCCATTCAGCTTCACCCAGCAGACATGCCAAACTCCGTCGTCA 1667
Qy 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
Db 1668 GATCTACAAGGAAGTGTGTACTGCAAACTCACTGTG 1704

RESULT 7
AY008764 2174 bp mRNA linear ROD 21-NOV-2000
LOCUS Mus musculus sentrin/SUMO-specific protease (SENp3) mRNA, complete
DEFINITION
ACCESSION AY008764
VERSION AY008764.1 GI:11245812
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL 1 (bases 1 to 2174)
MEDLINE Yeh,E.T., Gong,L. and Kamitani,T.
20267842 Ubiquitin-like proteins: new wines in new bottles
1-14 (2000)
PUBMED 10806345
REFERENCE 2 (bases 1 to 2174)
AUTHORS Gong,L. and Yeh,E.T.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2000) Institute of Molecular Medicine, University
of Texas-Houston, 2121 W. Holcombe Blvd. 928, Houston, TX 77030,
USA
FEATURES
source Location/Qualifiers
1..2174
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/chromosome="11"
1..2174
/gene="SENp3"
52..1758
/gene="SENp3"
/codon_start=1
/product="sentrin/SUMO-specific protease"
/protein_id="AAG33253.1"

```

```

/db_xref="GI:11245813"
/translation="MKETIOGTGSGWPGPGGTTYSNPRRERLRWLPKPKRUKSGS
GFPGDPGSGTIVPTRLPAPRPSFADASAEDEEEDDEEVAARLPPRWGLGAS
QRSRALRPSHRKTCQRRARRAFQMLLYSKSLTFHWKLMGRHRRKRLAHPKN
HLSPEGGATQVPSPCCRDFSPRLGPPRLGLGALMAEDMGSGSPVPSGPPMEED
GLRWTPKSLDPDPSGLSCTLPNGFGLSGPGEGRSLAPDASLILSNVCSIGDRVAQ
ELFOSSDLGIAEADRTGKAGQHSPLREBHVTCVQSILDFLQTYGSLILPLSTDDEV
EKLEDI FQOEFSTPSRKSLVQLQIQQSYORMPGNAMVRFVSYKRHVLTMDLCTLYG
ONWLNDOVMNMGDLVMDTPEKVFHFFNSFYDKLRTKGYDGVKRWTKNVDIFNKELL
LPIHLVHWSLISVDVRRRTIIFYDSQRLNRRCPKHIKAYLQAEAVKRDRLDFHOG
WKGYGFNMNVARQNNDSDCGAFVIOGKHLALSQPFSTQDMPKLRQIYKELCHCKL
TV"
BASE COUNT 513 a 585 c 570 g 506 t
ORIGIN
Alignment Scores:
Pred. No.: 3.22e-159 Length: 2174
Score: 1696.00 Matches: 320
Percent Similarity: 97.60% Conservative: 5
Best Local Similarity: 96.10% Mismatches: 7
Query Match: 95.71% Indels: 1
DB: 10 Gaps: 0
US-09-848-852A-3 (1-332) x AY008764 (1-2174)
Qy 1 MetTyrSerAlaGlnArgPheTyrGlyThrIleTyrAlaArgGlyAlaHis-LeuAl 20
Db 759 ATGTACTCTACCAATGTTTGGGGACTGTCTGGTCTCTGAAGGGAGCCGAGCTGGC 818
Qy 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
Db 819 ACCCCTGATGCCAGCATCTCATCAGTAATGTGTGCAGCATTTGGAGACCAGCTGGCTCA 878
Qy 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60
Db 879 GGAACCTTTTCAGAGCTCTGACTTGGGCATTGCGAAGAGGAGCATCGAGCTGGGGAGAA 938
Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 939 AGCTGGCCAGCATAGCCCTTCGGGAGGAGACATGTGACCTGTGTGCAGAGTATCTTAGA 998
Qy 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValAlaGluLy 100
Db 999 TGAATTCTCTCAAACTTATGCGAGCTCATCCCTCTCAGCACTGATGAAGTTGTAGAGAA 1058
Qy 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGl 120
Db 1059 GTTGAGGACATTTTCCAGCAGGAGTTCTCTACACCCCTCAAGGAAGAGCTCTGGTACTACA 1118
Qy 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 1119 GCTGATCCAGCTTATATCAGCGGATGCCAGGCAACGCTATGTTAGGGGCTTCGGGGTATC 1178
Qy 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160
Db 1179 CTATAAGCGACACGCTGCTCACCATGGATGACTTGGGTACCTTATATGACAGCAACTGGCT 1238
Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
Db 1239 CAATGACCAAGGTGATGAACATGTATGGAGATCTGGTCTATGGACACAGTCCCTCGAAGAGGT 1298
Qy 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
Db 1299 GCATTCTTCAACAGCTTCTTCTATGATGATAAATCCGTACCAAGGGTTATGATGGGGTAAA 1358
Qy 200 sArgTTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220
Db 1359 GAGGTGGACCAAAAATGTGGACATCTTCAATAAGGAATTAAGTCTTAATCCCATCATCT 1418
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
Db 1419 GGAGGTGCACCTGGTCCCTTATCTCAGTTGATGTAGGAGGAGGTACCATCACCTATTTTGA 1478

```

```

QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaG1 260
|||||
Db 1479 CTCCAGCAACTCTAAATCCCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAGA 1538
|||||
QY 260 uAlaValLysAspArgLeuAspPheHisGlnGlyTyrLysGlyTyrPheLysMetas 280
|||||
Db 1539 GGCAAGTCAAAAAAGACCGACTGACTTCCATCAGGCTGGAAGGGTACTTCAAAATGAA 1598
|||||
QY 280 nAlaIleArgGlnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300
|||||
Db 1599 TGTGGCCAGCAGCAATAATGACAGTACTGTGTGCTTTGTATTACAGTACTGCAGCA 1658
|||||
QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgG1 320
|||||
Db 1659 CCTGGCCCTGCTCAGCAATTCAGCTTACCCAGCAGGACATGCCCAAACTCGTGTCA 1718
|||||
QY 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
|||||
Db 1719 GATCTACAGGAAGTGTGTCTACTGCAACTCACTGTG 1755
|||||
RESULT 8
AC126239
LOCUS
DEFINITION
AC126239
VERSION
AC126239.1 GI:21693903
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
cat.
ORGANISM
Felis catus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
1 (bases 1 to 139405)
Benjamin,B., Blakesley,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Latic,P.,
Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masello,C., Maskeri,B., Mastrian,S.D.,
McCluskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
Portnov,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 139405)
Green,E.D.
Direct Submission
Submitted (04-JUL-2002) NIH Intramural Sequencing Center, 8717
Grosvmont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: daa
Center clone name: 474H17
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 131028 bases at least Q40
Consensus quality: 132577 bases at least Q30
Consensus quality: 133578 bases at least Q20
Insert size: 126000; agarose-fp
Insert size: 138305; sum-of-contigs
Quality coverage: 7.22x in Q20 bases; agarose-fp
Quality coverage: 6.57x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2527: contig of 2527 bp in length
2528 2627: gap of unknown length
2628 5297: contig of 2670 bp in length
5298 5397: gap of unknown length
5398 7601: contig of 2204 bp in length
7602 7701: gap of unknown length
7702 9908: contig of 2207 bp in length
9909 10008: gap of unknown length
10009 12523: contig of 2515 bp in length
12524 16221: contig of unknown length
16222 16321: gap of unknown length
16322 24505: contig of 8184 bp in length
24506 24605: gap of unknown length
24606 35941: contig of 11336 bp in length
35942 36041: gap of unknown length
36042 47449: contig of 11408 bp in length
47450 47550: gap of unknown length
47551 66344: contig of 18694 bp in length
66345 101268: contig of 34925 bp in length
101269 101368: gap of unknown length
101369 139405: contig of 38037 bp in length.
Location/Qualifiers
1..139405
/organism="Felis catus"
/db_xref="taxon:9685"
/clone="RP86-474H17"
/clone_lib="RP86"
1..2527
/note="assembly_fragment"
2628..5297
/note="assembly_fragment"
5398..7601
/note="assembly_fragment"
clone_end:Sp6
vector_side:left
7702..9908
/note="assembly_fragment"
10009..12523
/note="assembly_fragment"
12624..16221
/note="assembly_fragment"
16322..24505
/note="assembly_fragment"
24606..35941
/note="assembly_fragment"
36042..47449
/note="assembly_fragment"
47550..66243
/note="assembly_fragment"
66344..101268
/note="assembly_fragment"
101369..139405
/note="assembly_fragment"
BASE COUNT 32325 a 36788 c 37913 g 31272 t 1107 others
ORIGIN
Alignment Scores:
Pred. No.: 3,07e-84 Length: 139405
Score: 969.50 Matches: 320
Percent Similarity: 22.51% Conservative: 3
Best Local Similarity: 22.30% Mismatches: 9
Query Match: 54.71% Indels: 1107
DB: 2 Gaps: 8
US-09-848-852A-3 (1-332) x AC126239 (1-139405)

```


Db 31726 TTAAGCAGGCTGGGAAGAGCCTCCTCTGTTTGTACGGAGATTGGAAGTGGGCAAGAGTGGG 31785
QY 205 ----- 205
Db 31786 AGCAGGGGACTAACTAGGAGCGAGCAAGAGGAGGCGAGACACGCTCGACAGGATGC 31845
QY 205 ----- 205
Db 31846 TGAGGGAGGAGCAGGAGCGCTCTCTAGGATGGGTTTTTGTGAGGTAGAACCGAAGTACTTTC 31905
QY 205 ----- 205
Db 31906 TGGAGGGTTGGACAGAGGAATGAAGAAAACTCGGAATCAGGCTGATGCTGGGTTTGTGA 31965
QY 205 ----- 205
Db 31966 GCAGACGTTACTGGGGAAGTAGAGAGACCACTTTTGGCGAGGGGAAAGATAGTTCT 32025
QY 205 ----- 205
Db 32026 CGTTAGTTTCCCTCGCCTGCAATTCGAAGGCCAGGAAATTTCTACAGTGAGGTAGGACCG 32085
QY 205 ----- 205
Db 32086 TGGGCTGCAGATTTCTTAGTTTCTGTCCCGAGGCTCAGGAGCTCACACCTCCCGACC 32145
QY 205 ----- 205
Db 32146 TGGCAGGCTGTCCCGATAGAGGTGGGGGACCTACAGTGAAGAACTTGGCCTCTGCA 32205
QY 205 ----- 205
Db 32206 TACTGCATCTGAGTCCCACTCTGCCCTTCCGTGACCTTCAGCAAGTTGCCCTAGCTTC 32265
QY 205 ----- 205
Db 32266 TAAAGTGCCTTTTCCTCCGTTAGATCAGGAGGGAACCTTCTCGGAGGATTGTTGTGT 32325
QY 205 ----- 205
Db 32326 AACTGACCGACAGTTGCCGGGACCCAGGGTGTCTCCTCGGTGTGACTTAATTTCTCAGAA 32385
QY 205 ----- 205
Db 32386 CAAACAGAGGAGTAGCGGTGTCGTTTCCAGTGAAGAGAACGGGCGTATAGAGTGC 32445
QY 205 ----- 205
Db 32446 TCGCCCATGCCCCAGAGCTAGGAAAGGAGAGCTAGATCTAAACCCCCCAGTCTGGCT 32505
QY 205 ----- 205
Db 32506 CCAGAGTGTGACCTCGGTTCTTCCCTCTGAGCACTGGTACCTTTAATGTGTAGAAAG 32565
QY 205 ----- 205
Db 32566 TACCCGGGCATCTCGGCTCTCAGGTGTGTCTGTACACCCGCCCTTCCGGTGGAGG 32625
QY 205 ----- 205
Db 32626 CTGGGCTTTGTACCTTCCATCGGGCTTCTCGCCAGCGCTCCACTCGTCTCTAGACCCC 32685
QY 205 ----- 205
Db 32686 TCGTGAGGTAGCGGGACTCTCGTGCCCTCTGACAGACCCCTGGCACTGCCCTTTA 32745
QY 206 -ValAspIlePheAsnLysGluLeuLeuLeuLeuProIleHisLeuGluValHisTrpSe 225
Db 32746 GGTGGACATCTTCAATAAGAGGACTACTGTGATCCCATCCACCTCGAGGTGCATTGGTC 32805
QY 225 rIeuLleSerValAspValArgArgThrIleThrPheAspSerGlnArgThrIle 245
Db 32806 CCTCATCTCGGTTGAGTGAGGCGGACCATCATCACTACTTTGACTCGCAGCGCACCT 32865

QY 245 uAsnArgArgCysProLys----- 251
Db 32866 GAACCCCGCTGCCCTAA-CGTCTGAGTGGGAGGAGGCGGGAAGAGTGGGAGGA 32924
QY 251 ----- 251
Db 32925 AATCCAGTGGCGGCACTCGAGGCGGAAGGTGGGCTCTCTGAGGAGCAACCTG 32984
QY 251 ----- 251
Db 32985 GCGTTGGTCCCTGTGACCACCATACCCCGGCTCTCTGTGGGAGACACCGGGCATCCTCC 33044
QY 252 -----HisIleAlaLysTyrLeuGlnAlaGluAlaValLysLys 264
Db 33045 CTCCCTTTCTCCCTCTGCACAGCATATTCGCAAGTACCTACAGGCAGAGCGCTGAAGA 33104
QY 264 sAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMet----- 279
Db 33105 AGACCGCTGGATTTCACACAGGCTGGAAGGTTACTTCAAAATGGTAAGTTCTTAGAG 33164
QY 279 ----- 279
Db 33165 GGAGGAGTACGGGTGGTGACAGAGGTAAAGCAGGAGGAGAGTGTGAGGCCACACCCCT 33224
QY 280 -----AsnValAlaArgGlnAs 285
Db 33225 AGGACCTCCAAACGNAAGGCCCTCTCTCTCTCTTTCTAGAAATGTGCCAGGCAGAA 33284
QY 285 nAsnAspSerAspCysGlyAlaPheValLeu----- 295
Db 33285 TAATGACAGTACTGCGGTGCCTTTGTGTACAGGTAAGCAGATGCGAGCTGGGTTAAAG 33344
QY 295 ----- 295
Db 33345 GCGTGGCAGGAGGAGGAGGCTGTCTCAACCCCGCCCGCGTTGCTCAGAGCCC 33404
QY 295 ----- 295
Db 33405 CTTGTGGCAGAGAGTAACCTTAGGTGTCAAGAACGCAACATGGATTTCGTGGCTC 33464
QY 295 ----- 295
Db 33465 TCCTGGCTTCCCGTCCACATCGGACTGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 33524
QY 295 ----- 295
Db 33525 GAAGAGTGGCTTCTTAACCGCAATGCAGAGTATTTTAACACCGGATTTTCAGGTACCA 33584
QY 295 ----- 295
Db 33585 AGAGCAGGATTTTAGACTTTGAGCCCCAGGCTGGGGCACCTTGGCTCCTTTGTGAAT 33644
QY 295 ----- 295
Db 33645 GTGGATGTTCTCTGAAAGTTCCAGGAGGTGGATCTCACTCACAGAGGGGTAGCACA 33704
QY 296 -----GlnTyrCysL 299
Db 33705 GGTCTGACTTGGGGGTGGGGATGCTCACGCCCTGCTTTGTTAACACCCAGTACTGCA 33764
QY 299 yHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgA 319
Db 33765 AGCAGCTGGCGTGGCCAGCGGTTTCAGCTTCAACCAGGAGACATGCCAAACTTCGTC 33824
QY 319 rgGlnIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
Db 33825 GGCAGATCTACAAGAGCTGTGTCACTGCAAACTCACTGTG 33865

RESULT 9

AC126925/c

LOCUS

DEFINITION

AC126925

161428 bp

DNA

linear

HTG 10-JUL-2002

Canis familiaris clone RP81-332E11, WORKING DRAFT SEQUENCE, 15

Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysVal----- 75
|||||
Db 87138 GGCTGGCCAGCACAGCCCTCGGGAGGAGCATGTGACCTGCTACAGAGTAAGGACC 87079
Qy 75 ----- 75
Db 87078 CTAGAGACCCAGTCTCTCTAGTAGCCAACTCTTAAGTTGGAATCTGGAGCCCTCAC 87019
Qy 76 -----GlnSerIleLeuAspGluPheLeuGlnThrTyrGlyse 88
|||||
Db 87018 GCCACTCTCTTCCCTCCCTCAGGC-ATCTTGGATGAATTCCTTCAAACTATATGGCAG 86960
Qy 88 rLeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGlnG 108
|||||
Db 86959 CCTCATCCCTCTTAGCACTGATGAGGTGGTAGAGAAATTGGAGGACATTTTTTCAGCAGGA 86900
Qy 108 uPheSerThrProSer----- 113
|||||
Db 86899 GTTCTCTACACCTTCCAGGTGAGGCATGAAGAGAGATGTCGTGAGAGGGCTGGGCCCT 86840
Qy 113 ----- 113
Db 86839 GAGAGTATGAGAGGTTCGTGCTGCTTTCTCCCATAGGGATGTACAAGGCGGATGGAG 86780
Qy 113 ----- 113
Db 86779 GAAGGTAGAACTGAGGGGAGGAACCTGTAGGCAGAGTGCCACGCCCTGTAAACTGACT 86720
Qy 113 ----- 113
Db 86719 GGGGTGGAGTCTTACCTGAGATCCTGCAATATTATGATATACCTGATACCTCTGTGCCTC 86660
Qy 113 ----- 113
Db 86659 TCTGGAGACTAGGCCTTTACAGGTCCTCAAAAGGATGTGTGCCCAAGAAAGGATTAG 86600
Qy 113 ----- 113
Db 86599 TTTTGTGTTGAAGTGGTTCACACCCAGTGTGAAGCCCAACACAGGGCTTGAACCTCAG 86540
Qy 113 ----- 113
Db 86539 ACCTTGAGATCAATACCTGAGCTGAGATCAGGAATCGGATGTTTAAACCGACTGGGCCACC 86480
Qy 113 ----- 113
Db 86479 TAGTGGCCCCACCCAGGTGCCCTTATCTTAAGAATAAGAACTAGTTTAGGTCTT 86420
Qy 113 ----- 113
Db 86419 ATTCTGGCAGTAGAAAAGTACTTCTGTATGCTCTAGCCCCCTTGTCTCTGTGTGACTC 86360
Qy 114 -----ArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMe 128
|||||
Db 86359 CACCTTGGCCTACTCAGGAAGGCCGTGTGCTGACGTGATCCAGTCTGATCCAGCGGAT 86300
Qy 128 tProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThrMe 148
|||||
Db 86299 GCCAGCAATGCCATGGTGAGGGGTTTCCGAGTGGCCTATAAGCGGCATGTACTGACCAT 86240
Qy 148 tAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAsp----- 162
|||||
Db 86239 GGATGACCTGGGACCTTGTATGGACAGAACTGGCTCAATGACCAGGTGAGGAGTAGTGG 86180
Qy 162 ----- 162
Db 86179 AGAAATGGCCCCAAAAGGGGATTTAGGGAGCAGGGTGTCTGGGGCCCTCTGTTGGGGGA 86120
Qy 163 -----GlnValMe 165
|||||
Db 86119 GCCTATACCTATGCTGCACTCTCCATGGCAAGTGGCTCCCAATTTCTCCCGAGGTGAT 86060
Qy 165 tAsnMetTyrGlyAspLeuValMetAspThrValProGluLys----- 179

Db 86059 GAACATGTATGGAGACCTGTGTCTATGGACAGCTCCCTGAAAGGTAGGCCCAACACGAGT 86000
Qy 179 ----- 179
Db 85999 ACCTCAGTCCCCAGAAAGCCCTTCTGCAGTTTGAAGCAGCTTTTTCAGTCCCCCTTATCTC 85940
Qy 179 ----- 179
Db 85939 TTCATCCATAAAGAGATCTTGGTTTGGGGGAGAGGTGGTAGAAGGCAGTCACTTAAT 85880
Qy 179 ----- 179
Db 85879 CTGTAATCTTCTCTTTTAAAGTTTCTTATTTGAGAGAGAAAGGAGGGGGGAGG 85820
Qy 179 ----- 179
Db 85819 GAGGGAGAGTGGGGGGAAGCAGGAAATCTCCAGCAGATTCCCTGCTGAGTGCAGAGCC 85760
Qy 179 ----- 179
Db 85759 CAACGAGGGCTTGAGCTCTCAATACCGAGGGCTTGAGCTCTCAATACCGAGATCACAA 85700
Qy 179 ----- 179
Db 85699 CCCACGCCGAACCAAGAGTCAAGACTCAACTGAGTCACTGAGTCACTGAGTCACTCAAA 85640
Qy 179 ----- 179
Db 85639 TTTGTAATCGTGTCTGAACTTACTTAATCTTTTCCCGCAGTGTAGGGTTTTTCCCTGT 85580
Qy 179 ----- 179
Db 85579 GGTAACTCCAAATTTGGATGGGAGCAGTAAGATAAGTTAAAGAGCTTTGTTGTTT 85520
Qy 179 ----- 179
Db 85519 TTACCCTCAGTGTGAGTGCAGGTATTGAAACTGGCCCTTGGAGCCCTAATGAAAG 85460
Qy 179 ----- 179
Db 85459 GAAGCAAGCGCCAGGTGGAGGGGACCTGCAAGGGGCCCTGAGCTCTTTCAGTTCCAT 85400
Qy 180 -----ValHisPhePheAsnSerPhePheTyrAspLysL 191
|||||
Db 85399 CCAGGTCTTTTGTATCATTTGGCACAGGTGCATTTCTTCAACAGTTTCTTCTATGATAAAC 85340
Qy 191 euArgThrLysGlyTyrAspGlyValLysArgTrpThrLysAsn----- 205
|||||
Db 85339 TCCGCACCAAGGTATGATGGGTGAAAGGTGGACCAAAACAGTGAGTTATCAATTCA 85280
Qy 205 ----- 205
Db 85279 TATACGCTGGTGTGACACCTTGCTCTGAGAGGATTAGAGTTCTGTGTTCTTAGGAGCCCTT 85220
Qy 205 ----- 205
Db 85219 CCTAAAGCCCTGCTCACTCAGTCTTTCAAGTATTATAAGGCTCTGGTATGTGTGAGGTA 85160
Qy 205 ----- 205
Db 85159 GTCTAGGCACAAGGATTCACTGTTGAGTAAGGAAACCTTGCCCTTTATGGAGTTAAGGG 85100
Qy 205 ----- 205
Db 85099 TAAATAAATAGGTTAAATTTTCAGCTAGTGTAGTGTCTATGAATCACTCAAAAGGTAAT 85040
Qy 205 ----- 205
Db 85039 GCAGTAGCCCCGGAGCTAGGCCCTTTAGATTGGGGGTTTGGGAGGCCCTCTCTCTCTG 84980
Qy 205 ----- 205

Db	84979	AGGTTTGAAGCCCAAGGAGCTGGCCCTGCAGATCTGGGGGAGGAGTGGTCTAGGCAAAAG	84920
Qy	205	-----	205
Db	84919	AGTAGCAAGTACATGGGACCCAGAGTCAGAATGAGCATTTTCAGGCCACCGAAGGCCCATGT	84860
Qy	205	-----	205
Db	84859	GGGTGCTCTGTATAGTTGGAGCATCTGAGTGGAGGAGTGTAGTAAGAAATCCGGTCAGA	84800
Qy	205	-----	205
Db	84799	GTTGATCCTGTAAGGCATGGAGGAGCGCAGAGGAGATCTTCTTGATGTGCATGGAGA	84740
Qy	205	-----	205
Db	84739	AGCCACTGAAAAGTTTAAAGGAGCCTGGGAAGTCCTCATCTTTTCAACTCCTGCTTCAT	84680
Qy	205	-----	205
Db	84679	GTAGACTGGAAGTGGGGGGAAGAGTAGAAGCAGCAGACTGGCCAGGAGGGGAGCACAG	84620
Qy	205	-----	205
Db	84619	AAGCCTGGACGAGGGCGCTGAGTCCAGAAATGGAAGTGGCTTTCTCTAGGATAGGTTTTG	84560
Qy	205	-----	205
Db	84559	TGGGTAGAACCATGGTATTTTCAGATGGTGTGGATGTGGGANTGAAGNAGCTAGGANT	84500
Qy	205	-----	205
Db	84499	TGGGAGGGCCTGGTTAGTGGCAGAGACTTACTGGGAAAANTGAGGGAACGCGTTTGCTGA	84440
Qy	205	-----	205
Db	84439	TGGGTAAGGGTGGGTATCTTTTCTCCCTCTAGTTGGGGTCCCTGCCCCCAAG	84380
Qy	205	-----	205
Db	84379	GTGAAAAATTTGGTCTAGAGTGAAGTGGGACTGCCTCTGGCATTTTAGTTCTGTCCCCG	84320
Qy	205	-----	205
Db	84319	GGTTCAGGGGCTTACACCTGCAGTGACCTGSCAGGCCCGCTTCTCCAGTATGGTGGGA	84260
Qy	205	-----	205
Db	84259	TAACACAGTGGAAAGAACGTGGCTGTCCCCCTGTCTCCTGTCTCCTGTGCCCGTCAGA	84200
Qy	205	-----	205
Db	84199	AAGGTGCCCTACTCTCCAGGCTCCAGTTCATCGTTAGGCTGACTGAGAACTCCCTCAC	84140
Qy	205	-----	205
Db	84139	AGGCTGTTGTGTAAACGGCGCACCCCTTACTAGTCTGTGCGGTGCTTTAACTGAT	84080
Qy	205	-----	205
Db	84079	CATGGACAGGAACTAATTTAACTCTGGTAACAAAGTGAAGATAGGACTGTTATTTCGC	84020
Qy	205	-----	205
Db	84019	AGCCTCTTACATCTAAGAGACGGTGGCCACGAGAGCCATTCCGTAGCTACCCGTTGGTC	83960
Qy	205	-----	205
Db	83959	ACAACTAGGAATGAGGAGCTGGGTTTAAACCCAGCGGCTGGCTCCAGAACTCCTGA	83900
Qy	205	-----	205
Db	83899	GCCTGGTTTTTATGCTAATAGTACTCGTGCCTTTATGTGTGAAAGAAAGTACCTGGC	83840

QY	205	-----	205
Db	83839	CATCCTGGGTCTCAGGTGTGTGTGTGACAGTAGGTCTTCTGACCTGCCTTCTTCCCTCAC	83780
QY	205	-----	205
Db	83779	CCCTGGAATCTACACATGCTGAATCTCACCTCAGTGGAGGAGCTCAGTGTGTGTGTTG	83720
QY	205	-----	205
Db	83719	CCCACCTGGCTCCTTGTCTCCGCCTCCACTCTTTAGGACCTCCCGCAATGGCGCAGGTG	83660
QY	206	-----	211
Db	83659	GCTCTGCTGCTCAACCTGACCTCTGACACACACCTCTCTTAGGTGGACATCTTCAATAAG	83600
QY	212	GlulLeuLeuLeuIleProIleHisLeuCluValHisTrpSerLeuIleSerValAspVal	231
Db	83599	GAGCTCCTGCTAATCCCATCCACTCGAGGTGCACTGGTCCCTCATCTCTGTGTGACGTG	83540
QY	232	ArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgCysProLys	251
Db	83539	AGGCGACCACTACCTACTTTGACTCGCAGCGACCTTAACCGCGCTGCCCTAA-	83481
QY	251	-----	251
Db	83480	GGTTTGAGAGGAAGCGGGGGGGGTGGGGGGGTGGAGATCCAGTGGCAAGCAT	83421
QY	251	-----	251
Db	83420	AGTGGGAGAAGGCGAGGCGCTGTCCTGGGACCACTGTACCCTAGGTTTCACTTTGGGAAA	83361
QY	252	-----	259
Db	83360	CGAGGCGATCACTTCCTCTTTTCTCTCCATAGCATATGCCAAGTACCTACAGGCA	83301
QY	260	GlulAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMet	279
Db	83300	GAAGCAGTGAAGAAGACCGCTGATTTCCACCGAGGCTGGAAGGTTACTTCAAAATG	83241
QY	279	-----	279
Db	83240	GTAAGTGTCTGGAGGAGGGGTACAGGTGTGTATGTTGGTGTAGAGAGCGGTAGGAGG	83181
QY	279	-----	279
Db	83180	AAGAAGTGGAGCCACACCCGAGGAGGCTCCAGGTGAAGACTCTTTTTATTCCTA	83121
QY	280	-AsnValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeu	295
Db	83120	GAACGTGCCAGGCAGATATATGACAGTACTGTGTGGTCCCTTGTGTACAGGTAAAGCAG	83061
QY	295	-----	295
Db	83060	AGATCGGTGGGTAAAGGTTTGGTAAAGGAGGTGGGAGGCTGTTCACCTCTCTGTGTTG	83001
QY	295	-----	295
Db	83000	GCTCACAGGCATTTGTGAAGCAGAGAACCTTAGGTGTTAGTGTGCAGAAATCCGAT	82941
QY	295	-----	295
Db	82940	CCCCTTGTCTCCCGCTTCCCGGGGCCACACTTGGCCTGGGTCTCTTGGTCTTGGTC	82881
QY	295	-----	295
Db	82880	TCTGGCCCTGCTACAGTTGCTTCCCAACCCCAAGTGAAGCAATTTTCTCAGTATCGA	82821
QY	295	-----	295
Db	82820	GAACAGTGACTCTAGAGTCAAGCTTTAGGCTATGGAGCTTGGCTTTGTGTGTGT	82761

```

Qy 295 ----- 295
Db 82760 GAAATGTGGATGTTCTGTGAAAGTTTGGTGAGTGATGGCTCCTGTGCAGCACAGC 82701
Qy 295 ----- 295
Db 82700 ACAGGTCTGACCTGGGGGATGCTCAGTAGATGGTTCTTATGCGCTGCTTTGTTAAAC 82641
Qy 296 -GlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetPr 315
|||||
Db 82640 CCAGTACTGACGACCTGGCCCTGCTCAGCATTCAGTTCCACCGAGGACATGCC 82581
|||||
Qy 315 oLysLeuArgGlnIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
|||||
Db 82580 CAAACTTCGTGGCAGATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTG 82529
|||||

RESULT 10
AC126921/c
LOCUS
DEFINITION Bos taurus clone RP42-45D24, WORKING DRAFT SEQUENCE, 12 unordered
pieces.
ACCESSION AC126921
VERSION AC126921.1 GI:21724098
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 153553)
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C.,
Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J.,
Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P.,
Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,
Margulies, E.H., Masello, C., Maskeri, B., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Paguirigan, C., Pearson, R.,
Portnoy, M.E., Prasad, A., Schueler, M.G., Stantrop, S., Thomas, J.W.,
Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 153553)
Green, E.D.
Direct Submission
Submitted (10-JUL-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: ddi
Center clone name: 045D24
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 146066 bases at least Q40
Consensus quality: 147748 bases at least Q30
Consensus quality: 148824 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 152453; sum-of-contigs
Quality coverage: 8.80x in Q20 bases; agarose-fp
Quality coverage: 8.72x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

```

* as soon as it is available and the accession number will
* be preserved.
* 1
* 2313 2312: contig of 2312 bp in length
* 2413 2412: gap of unknown length
* 5841 5841: contig of 3429 bp in length
* 5842 5841: gap of unknown length
* 5942 8435: contig of 2494 bp in length
* 8436 8435: gap of unknown length
* 15799 15799: contig of 7264 bp in length
* 15800 15899: gap of unknown length
* 15900 25224: contig of 9325 bp in length
* 25225 25324: gap of unknown length
* 25325 32604: contig of 7180 bp in length
* 32605 32604: gap of unknown length
* 40970 40970: contig of 8366 bp in length
* 40971 41070: gap of unknown length
* 41071 56590: contig of 15520 bp in length
* 56591 56590: gap of unknown length
* 56591 73769: contig of 17079 bp in length
* 73770 73769: gap of unknown length
* 73770 90859: contig of 16990 bp in length
* 90860 90859: gap of unknown length
* 90860 111428: contig of 20469 bp in length
* 111429 111528: gap of unknown length
* 111529 153553: contig of 42025 bp in length.

FEATURES
Location/Qualifiers
Source
1..153553
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="RP42-45D24"
/clone_lib="RP42"
1..2312
/note="assembly_fragment"
2413..5841
/note="assembly_fragment"
vector_end:T7
vector_side:left"
5942..8435
/note="assembly_fragment"
8536..15799
/note="assembly_fragment"
15900..25224
/note="assembly_fragment"
25325..32504
/note="assembly_fragment"
32605..40970
/note="assembly_fragment"
41071..56590
/note="assembly_fragment"
56691..73769
/note="assembly_fragment"
73870..90859
/note="assembly_fragment"
90960..111428
/note="assembly_fragment"
111529..153553
/note="assembly_fragment"
vector_end:SP6
vector_side:right"

BASE COUNT 39178 a 37393 c 36616 g 39259 t 1107 others
ORIGIN

Alignment Scores:
Pred. No.: 3,23e-72 Length: 153553
Score: 849.50 Matches: 285
Percent Similarity: 22.21% Conservative: 4
Best Local Similarity: 21.91% Mismatches: 11
Query Match: 47.94% Indels: 1001
DB: 2 Gaps: 7

US-09-848-852a-3 (1-332) x AC126921 (1-153553)
Qy 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAla-HisLeuAl 20

```


FEATURES
source

Location/Qualifiers
1..1578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1001413"
/tissue_type="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stages="embryo, 10 weeks"
/note="cloning vector: pME18SFL3"
44..619
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91428.1"
/db_xref="gi:7021893"
/translation="MSRNSISIPVESLHVLMMLGSPFLGVGLTVLVDASVYPCGGL
ICMHSTWICPLCPQASLSCPFPICPIGLLSCTLPNGFGGSGPGEERSLAPDAS
ILISNVCSIGHVAQELFOGSDLGMAREAPRGKAGQHSPLREHVTVCVSIIDFL
QTYGLSLPLSTDEVKEKLEIDIFQEFFSPSR"
BASE COUNT 339 a 372 c 439 g 428 t
ORIGIN

Alignment Scores:
Pred. No.: 4,25e-72 Length: 1578
Score: 820.00 Matches: 186
Percent Similarity: 50.00% Conservative: 10
Best Local Similarity: 47.45% Mismatches: 17
Query Match: 46.28% Indels: 179
DB: 9 Gaps: 4

US-09-848-852A-3 (1-332) x AK000923 (1-1578)

QY 1 MetTyrSerAlaGlnArgPheThrGlyThrIleTpAlaArgArgGlyAlaHis-LeuAl 20
|||||
DB 274 ATGTACTCTGCCACAGGTTTGGGGACATCTGGGCACAGAGGGAGCGGAGCTTGGC 333
QY 20 aProProaspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
|||||
DB 334 ACCCCCTGATGCCAGCATCTCATCAGCAATGTGTGCAGCATCGGGGACCATGTGGCCCA 393
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60
|||||
DB 394 GGAGCTTTTTCAGGCTCAGATTTGGGGATGGCAGAGGAGGAGGAGGAGGAGGAGGAGAA 453
QY 60 sAlaGlyGlnHisSerProLeuArgGluHisValThrCysValGlnSerIleLeuAs 80
|||||
DB 454 AGCCGCGCCAGCACACCCCTCGAGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTGA 513
QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100
|||||
DB 514 CGAATTCCTTCAACGATATGGCAGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGAA 573
QY 100 sLeuGluaspIlePheGlnGlnPheSerThrProSer----- 113
DB 574 GCTGGAGGACATTTTCCAGCAGGAGTCTTCCACCCCTTCCAGGTCAGGCTTGAAGCCCT 633
QY 113 ----- 113
DB 634 CTTTCAAGAGAGGCTGGGGCTTGGGATGTGGAGAGAACTACTGCTGCTTTCTTCC 693
QY 113 ----- 113
DB 694 TAGGGCTGTAGTTGGGAGGAGGAAGCTAGAGCTGAAGGGGAGGACTCTGCAGGCGAGGT 753
QY 113 ----- 113
DB 754 CCAATTCCTTGAAGCTGATGGGAGAGTCTTTACCTGGGACCCCTGAAATGTTCTACCTGA 813
QY 113 ----- 113
DB 814 GTAGTCATGTTTATCTTTCTGGGGAGTGGGCTTTTCGCGAGGTCTTCAGAGACCCATCA 873
QY 113 ----- 113

Db 874 TGAGCCAGAAAAAGGGAGCTCAGTTAGATTGTTATTCAGGGAGTAGTAGTATTTC 933
QY 114 -----ArgLysGlyLe 117
DB 934 TGTGCCCCAGCTGCATCATCTTTTGTGTGACTCCACCCTTGGCCCTACTCAGGAAGGGCCT 993
QY 117 uValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPh 137
|||||
DB 994 GGTGTTGCAGCTGATCCAGCTCTTACCAGCGGATGCCAGCAATGCCATGCTGTCAGGGCCT 1053
QY 137 eArgValAlaTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGl 157
|||||
DB 1054 CGAGTGGCTTATAAGCGGCACGTGCTGACCATGTGATGACTTTGGGACCTTGTATGGACA 1113
QY 157 nAsnTrpLeuAsnAsp----- 162
|||||
DB 1114 GAATCGCTCAGTGACCGAGGTAGAGAAAGGCTAGAGAACAGCCCTGAGAGGGGATTCAGG 1173
QY 162 ----- 162
DB 1174 GAGCAGGTGTCTGGGGCCCTCTGCATGGGGAGGCCCTGTACCCATGCCGACCCCTCCAT 1233
QY 163 -----GlnValMetAsnMetTyrGlyAspLeuValMetAs 174
|||||
DB 1234 GGCAAGCTGCCCTCCCATCTTCTCCCGAGGTGATGAACATGTATGGAGACCTGTGCATGA 1293
QY 174 pThrValProGluLysVal-----HisPhePheAsnSe 185
|||||
DB 1294 CACAGTCCCTGAAAGGTAGGCCCAACCAGATAGTCTACCCAGAGAACTTCTGCAG 1353
QY 185 rPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTyrThrLysAs 205
|||||
DB 1354 TTTT-----AGCAGCTTTTAAAGCTCTTTCATCTCTTTCATTACACAGAGAG 1404
QY 205 nValAspIlePheAsnLysGluLeuLeuLeu 216
|||||
DB 1405 GGTCTCTGTTTCAGGAGAGAGGTGTAGTGTA 1438
RESULT 12
AC099294
LOCUS
DEFINITION Rattus norvegicus clone CH230-39c15, *** SEQUENCING IN PROGRESS
***, 71 unordered pieces.
AC099294
VERSION AC099294.4 GI:21731823
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 257644)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Loz do, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.
 Direct Submission
 2 (bases 1 to 257644)
 Worley, K.C.
 Direct Submission
 Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 257644)
 Worley, K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 11, 2002 this sequence version replaced gi:20258090.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GUSH
 Center clone name: CH230-39C15
 ----- Summary Statistics
 Sequencing vector: plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 192798 bases at least Q40
 Consensus quality: 198124 bases at least Q30
 Consensus quality: 202420 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 71 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1665: contig of 1665 bp in length
 * 1666 1765: gap of unknown length
 * 1766 2950: contig of 1185 bp in length
 * 2951 3050: gap of unknown length
 * 3051 4076: contig of 1026 bp in length
 * 4077 4176: gap of unknown length
 * 4177 5352: contig of 1176 bp in length
 * 5353 5452: gap of unknown length
 * 5453 6762: contig of 1310 bp in length
 * 6763 6862: gap of unknown length
 * 6863 8152: contig of 1290 bp in length
 * 8153 8252: gap of unknown length
 * 8253 9323: contig of 1071 bp in length

9324 9423: gap of unknown length
 9424 10907: contig of 1484 bp in length
 10908 11007: gap of unknown length
 11008 12354: contig of 1347 bp in length
 12355 12454: gap of unknown length
 12455 14718: contig of 1718 bp in length
 14719 14722: gap of unknown length
 14723 15793: contig of 1521 bp in length
 15794 15894: gap of unknown length
 15895 17223: contig of 1230 bp in length
 17224 18364: contig of 1141 bp in length
 18365 18464: gap of unknown length
 18465 19003: contig of 1439 bp in length
 19004 20003: gap of unknown length
 20004 21256: contig of 1253 bp in length
 21257 21356: gap of unknown length
 21357 22431: contig of 1075 bp in length
 22432 22531: gap of unknown length
 22532 24064: contig of 1533 bp in length
 24065 24164: gap of unknown length
 24165 25804: contig of 1640 bp in length
 25805 25904: gap of unknown length
 25905 27691: contig of 1787 bp in length
 27692 27791: gap of unknown length
 27792 29442: contig of 1651 bp in length
 29443 29542: gap of unknown length
 29543 30720: contig of 1178 bp in length
 30721 30820: gap of unknown length
 30821 32491: contig of 1671 bp in length
 32492 32591: gap of unknown length
 32592 35298: contig of 2707 bp in length
 35299 35398: gap of unknown length
 35399 37535: contig of 2137 bp in length
 37536 37636: gap of unknown length
 37637 39829: contig of 2094 bp in length
 39830 43076: contig of 3247 bp in length
 43077 43176: gap of unknown length
 43177 46452: contig of 3276 bp in length
 46453 46552: gap of unknown length
 46553 48963: contig of 2411 bp in length
 48964 49063: gap of unknown length
 49064 51571: contig of 2508 bp in length
 51572 51671: gap of unknown length
 51672 54841: contig of 2970 bp in length
 54842 54741: gap of unknown length
 54742 58045: contig of 3304 bp in length
 58046 58145: gap of unknown length
 58146 61211: contig of 3066 bp in length
 61212 61311: gap of unknown length
 61312 63581: contig of 2270 bp in length
 63582 63681: gap of unknown length
 63682 66226: contig of 2545 bp in length
 66227 66326: gap of unknown length
 66327 68907: contig of 2581 bp in length
 68908 69007: gap of unknown length
 69009 73249: contig of 4242 bp in length
 73250 73349: gap of unknown length
 73350 77046: contig of 3697 bp in length
 77047 77146: gap of unknown length
 77147 80158: contig of 3012 bp in length
 80159 80258: gap of unknown length
 80259 82863: contig of 2605 bp in length
 82864 82963: gap of unknown length
 82964 86678: contig of 3715 bp in length
 86679 86778: gap of unknown length
 86779 90147: contig of 3369 bp in length
 90148 90247: gap of unknown length
 90248 94775: contig of 4528 bp in length
 94776 94875: gap of unknown length
 94876 97958: contig of 3083 bp in length
 97959 98058: gap of unknown length

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

*	98059	101384:	contig of 3326 bp in length	
*	101385	101484:	gap of unknown length	
*	101485	105449:	contig of 3965 bp in length	
*	105450	105450:	gap of unknown length	
*	105550	109216:	contig of 3667 bp in length	
*	109217	109316:	gap of unknown length	
*	109317	113575:	contig of 4259 bp in length	
*	113576	113675:	gap of unknown length	
*	113676	116968:	contig of 3293 bp in length	
*	116969	117068:	gap of unknown length	
*	117069	120956:	contig of 3888 bp in length	
*	120957	121056:	gap of unknown length	
*	121057	126698:	contig of 5642 bp in length	
*	126699	126798:	gap of unknown length	
*	126799	131591:	contig of 4793 bp in length	
*	131592	131691:	gap of unknown length	
*	131692	135524:	contig of 3833 bp in length	
*	135525	135624:	gap of unknown length	
*	135625	139656:	contig of 4032 bp in length	
*	139657	139756:	gap of unknown length	
*	139757	144725:	contig of 4969 bp in length	

Alignment Scores:

Pred. No.:	2.29e-66	Length:	257644
Score:	794.00	Matches:	164
Percent Similarity:	62.43%	Conservative:	62
Best Local Similarity:	45.30%	Mismatches:	89
Query Match:	44.81%	Indels:	48
DB:	2	Gaps:	6

US-09-848-852A-3 (1-332) x AC099294 (1-257644)

QY	11	IleTtPAlaAqArgGlyAlaHisLeuAlaProAspAlaSerIleLeuIleSer---	29
Db	205854	CTCTGGGCAGGAAGCGGACAACT-GCTGTCACCGATGACGGAGTGAAGCTGTCAGTG	205912
QY	30	-----AsnValCysSerIleGlyAspHisValAlaGlnLeuPheGln	44
Db	205913	TCTGGAGCAGACGAATCTGTGAGTAGTAGATGGCGCTGTGTCAGGAGCCGCTCAG	205972
QY	45	GlySerAspLeuGlyMetAlaGlu-----	52
Db	205973	AATAAGAACTTCCAGATGGAGGAGGATGGGTCTCTCAGGCAGAACATTTAGTTTGAAA	206032
QY	52	-----	52
Db	206033	CTGCTGGACCACCTTACTGTAAGAGTCACATGGATGCTCCCTGGTGTGACGTGAGCTC	206092
QY	53	-----GluAlaGluArgProGlyGluLysAlaGlyGln---HisSerProLeuArgGlu	69
Db	206093	AAAGTAGAAACCAAGTGGCAGGTGGCAAGAGTGCCAGACAGCGTCTCCCGTGGATGAT	206152
QY	70	GluHisVal--ThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySer	88
Db	206153	GAGCAGCTCTGCACCTGCTTCTGGGTCTCTAGACGAGGTATGAAAAAGTATGGAAGT	206212
QY	89	LeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGlnGlu	108
Db	206213	CTGGTCCCACTCAGTCAGAAAGATGCTCTGGGAGATTAAGACATGCTTTAATGAAGAC	206272
QY	109	PheSerThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMet	128
Db	206273	TTTTTCT-----AATAGAAAACCAATTTATCAATAGGAAATAACAACTACCGGCCAGA	206326
QY	129	ProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThrMet	148
Db	206327	CATCAAAAAGTGT-----AACTTCCGAATCTTCTACAAACACATGCTGGATATG	206377
QY	149	AspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyr	168
Db	206378	GACGACCTGGCAACTTTGGATGGTCAGAAATGGTTGAATGACCAAGTTTATTAATATGAT	206437
QY	169	GlyAspLeuValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyr	188

DB 206438	GGCGAGCTGATAAATGGATGCCAGTCCACAGCAAGGTCCACTTCTTCCACAGCTTTTTCAT	206497	
QY 189	AspLysLeuArgThrLysGlyTyrAspGlyValLysArgTyrTrpThrLysAsnValAspIle 208		
DB 206498	ACACAGCTGGTAACCAAGGCTATAATGGAGTTAAGACTGAGCTAAGAAGGTGGATTG	206557	
QY 209	PheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIleSer 228		
DB 206558	TTTAAAGAGTCTCTCTCTGATTCCCATCCACCTGGAGGTCCACTGGTCTCTCATTA	206617	
QY 229	ValAspValArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgArg 248		
DB 206618	GTGACACTCTCCAGTCGGATTATTTTATTTATGATTTCCCAAGGATTCATTTTAAATTT	206677	
QY 249	CysProLysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAsp 268		
DB 206678	TGTGTAGAGAAATATAAGAAAGTATTTGTCTGACTGAGCCAGAGAAAANAATAGACTGAA	206737	
QY 269	PheHisGlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAsnAspSer 288		
DB 206738	TTTCTTCAGGGTGGCAGACTGCTTTACAAAGTGTATTCCACACACAGAAAATGACAGT	206797	
QY 289	AspCysGlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPheSer 308		
DB 206798	GACTGTGGAGTCTTGTGCTCCAGTACTGCAAGTGCCTCGCCCTAGACAGCGCTTTCCAG	206857	
QY 309	PheThrGlnGlnAspMetProLysLeuArgArgGlnIleTyrLysGluLeuCysHisCys 328		
DB 206858	TTTTCTCAAGAAGACATCGCTCGAGTGGGAAGAGGATCTATTAAGGAGCTGTGTGAGTGC	206917	
QY 329	LysLeu 330		
DB 206918	CGGCTC 206923		
RESULT 13			
AB060892			
LOCUS			
DEFINITION	AB060892 2007 bp mRNA linear PRI 13-JUN-2001		
ACCESSION	AB060892		
VERSION	AB060892.1 GI:13874561		
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Macaca fascicularis adult male temporal lobe right CDNA to mRNA, clone.lib.macaque brain cDNA library Qtra clone:Qtra-13616.		
ORGANISM	Macaca fascicularis		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae; Cercopitheciinae; Macaca.		
AUTHORS	1 (sites)		
TITLE	Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.		
JOURNAL	Isolation of full-length cDNA clones from macaque brain cDNA libraries		
REFERENCE	2 (bases 1 to 2007)		
AUTHORS	Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.		
JOURNAL	Direct Submission		
TITLE	Submitted (27-APR-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan		
COMMENT	(E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/, Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)		
	Lab host: TOP10		
	Vector: pME18S-FL3 (Acc.No. AB009864)		
	R. Sitel: DraIII (CACTGTGTG)		
	R. Site2: DraIII (CACCATGTG)		
	Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside		

the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method (Sugano et al., , Institute of Medical Science, University of Tokyo).

Custom primer used for sequencing
(5' end primer [CTTCTGCTCTAAAGCTCGG];
3' end primer [CGACCTGCAGCTCGACACA]).

FEATURES

source
1. .2007
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="QtrA-13616"
/sex="male"
/tissue_type="temporal lobe right"
/clone_lib="maceque brain cDNA library QtrA"
/dev_stage="adult"
1140. .1442
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB46895.1"
/db_xref="GI:13874562"
/translation="WPCNAMVRGFRVAYKRHVLTMDDLGLYCONWLNDOVRKRGRTG
LRGDSRSRSGALCMGEPCHTHAAPSMASCLPSPQVMNMGDLVMDTVPEKVGPIR"
BASE COUNT 469 a 465 c 551 g 522 t
ORIGIN

Alignment Scores:
Pred. No.: 4, 65e-69 Length: 2007
Score: 791.00 Matches: 177
Percent Similarity: 50.86% Conservative: 1
Best Local Similarity: 50.57% Mismatches: 2
Query Match: 44.64% Indels: 170
DB: 9 Gaps: 2

US-09-848-852A-3 (1-332) x AB060892 (1-2007)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20
DB 382 GTTGACTCTGCCAACGGTTTGGGGACCACCTCTGGGCCAAGAGGGGAGCGAGCTGGC 441
QY 20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
DB 442 ACCCCCTGTATGCCAGCATCTCATCAACAATGTGTCCAGCATCGGGACCACATGTGCCCA 501
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60
DB 502 GGAGCTTTTCAGGCTCAGATTTGGGCATGGCAGAGAGGAGGAGGCTGGGGAGAA 561
QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
DB 562 AGCCGGCCACACAGCCCTCGCAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGA 621
QY 80 pGluPheLeuGlnThrTyrGlySerIleLeuProLeuSerThrAspGluValValGluLy 100
DB 622 CGAATTCTTCAACATATGCGACGCTCATACCCCTCAGCAGCTGATCAGGTAGTGGAGAA 681
QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSer- 113
DB 682 ATGGAGGACATTTTCAGCAGGAGGTTCACAGCCTTCACAGCTGAGGCTTGAAGCCCT 741
QY 113 ----- 113
DB 742 CCTTCAAGAGGGCTGGGGCCTCAGGAGGTAGGAGCTAGAGAGATACTGCTGCTT 801
QY 113 ----- 113
DB 802 TTCTCCATAGGCTGTACTTTGGGGAGGAGGAGGAGCTAGAGCTGAAGGGGAGGACTCTGC 861
QY 113 ----- 113
DB 862 AGGCAAGGTCGCCAGTCTTGGAGCTGATGGGAGAGTCTTTTACCTGGGACCCTGACATG 921
QY 113 ----- 113

Db 922 TTCTACTTGAGTAGTATCATGTTCTTCTTGGGCACTGGGCCTTTTCGGGGTCTCTCAGAAGGA 981
QY 113 ----- 113
Db 982 CCCATCATGAGCCGAATAAAAGGGGTCACTTAGAATTTGTATTCTAGGAGTAGTAGGT 1041
QY 114 -----ArgL 115
Db 1042 ATTTCTGTGCCCCAGCTGCATCAACTTTGTGTGCTGACTCCACCCTTGGCCTACTCAGGA 1101
QY 115 ysglyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetVala 135
Db 1102 AGAGCCTGGTGTTCAGCTGATCCAGTCATACCAAGGATGCCAGCAATGCCATGGTGA 1161
QY 135 rgGlyPheArgValAlaIaTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuT 155
Db 1162 GGGCTTCCGTGTGGCTATAGCGCATGTGCTGACCATGGATGACTTGGGACCTGT 1221
QY 155 yTgGlyGlnAsnTrpLeuAsnAsp- 162
Db 1222 ATGGACAGAACTGGCTCATGACCAGGTGAGAAAGGTTAGAGNAACAGCGCTGAGAGGG 1281
QY 162 ----- 162
Db 1282 ATTCAGGAGAGAGGTGTCTGGGGCCCTCTGTCATGGGGAGGCCCTGTACCCATGCCGCAC 1341
QY 163 -----GlnValMetAsnMetTyrGlyAspLeuV 172
Db 1342 CTCCATGGCAAGCTGCCCTCCCATCTTCTCCCGAGGTGATGAACATGTATGGAGACCTGG 1401
QY 172 aMetAspThrValProGluLysVal 180
Db 1402 TCATGGACACAGTCCCTGAAAGGTA 1427
RESULT 14
BC008589
LOCUS Homo sapiens, clone IMAGE:3448367, mRNA, partial cds.
DEFINITION BC008589
ACCESSION BC008589
VERSION BC008589.1 GI:14250318
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1830)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@bcm.tmc.edu.
Villalona, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 19 Row: e Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF

GRPVTNWRQLRHFGQRKKALQIOKTIWKDBHLCAKTKFNVAQNTVSLSSKVRKKDAK
HFTSSSKTLLRLQAOKLLSAAKSDHEYCREKNLLKAVTDPPSNAGQANGHRPRTD
PQSDPMPKNGESQSGESTIVTLLNNHKRGFCYGCQCGPEHNRNGGPLLKPKFO
LNOHRRITKLSPLMWKLSMIRYRITLRSOHERTKSKYCKLBAQSRVWQKYTGDOH
ETRENGEGSSCPFFSPKSDSCSRHQPIYFPDMSAVVYKNSHVPDCHTKGSSFL
GKELSDIAPDQOQCANAWQSSCSQWECTELIHDIPLPHERNDMTISETER
EIMTLQENGTSSVDNRKLVSGADTVSSVDGPVSQKAVNGNSYQMEEDGSLQK
SILSLELHDHPYCKSPLEAPLVCSGLKLENGVGGKNSOKASPVDDOELSVCLSGSLG
EVMKKYGSVLPSLSEKVEVLGRDKOVFNEDFSNRKPFINREITINRAHOKCNFRIFYNK
HMLDMDDLATLQDGNMLNDQINMYGELINDAVDPKVHFFNSFFHQLVTKYNGYKVR
WTKFKDLFKKSLLLIPIHLHVVHNSLITVTLNRIISFYDSQGHFFCFVENIRKYLIT
EAEKRNPEFLOGWQAVTKCIPOQKNDSDCGVFVLQYCKLCALEOPFQFSQEDMPRV
PKRTYKPIFCBTLMD

BASE COUNT	841 a	572 c	664 g	716 t
ORIGIN				

Alignment Scores:

Pred. No.:	1.18e-68	Length:	2793
Score:	789.00	Matches:	148
Percent Similarity:	73.7%	Conservative:	54
Best Local Similarity:	54.01%	Mismatches:	66
Query Match:	44.53%	Indels:	6
DB:	9	Gaps:	3

US-09-848-852A-3 (1-332) x BC030705 (1-2793)

Qy	58	GlyCluLysalaGlyGlnHisSerProLeuArgGluGluHisValThr--CysValGln	76
Db	1656	GAAGAACAACGATCAGAAAGCCCTCCAGCTGGATGATGAACAGCTGTCTAGCTGTCTTCT	1715
Qy	77	SerLeuLeuAspGluPheLeuGlnThrTyrglySerLeuLeuProLeuSerThrAspGlu	96
Db	1716	GGATTCCTAGATGAGGTATGGAAGAAGTATGGCAGTTGGTTCCACTCAGTGAAGAAGAA	1775
Qy	97	ValValGluLysLeuGluAspPheGlnGlnGluPheSerThrProSerArgLysGly	116
Db	1776	GTCTCTGGAAGATTAAAGATGCTTTAATGAAGACTTTTCT-----AATAGAAAAACCA	1829
Qy	117	LeuValLeuGlnLeuLeGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGly	136
Db	1830	TTTATCAATAGGGAATACAACTATCGGGCCAGACATCAAAATGT-----AAC	1880
Qy	137	PheArgValAlaTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGly	156
Db	1881	TTCCGTATCTTCTATAATAAACACATGCTGGATATGGACAGCTGGCGACTCTGGATGGT	1940
Qy	157	GlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrVal	176
Db	1941	CAGAAGCTGGCTGAATCACCAGGTCAATTATATGATGTGAGCTGATAATGGATGGATGCAGTC	2000
Qy	177	ProGluLysValHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyr	196
Db	2001	CCAGACAAAGTTCACCTCTTCAACAGCTTTTTCATACACAGCTGGTAACCAAGGATAT	2060
Qy	197	AspGlyValLysArgTrpThrLysAsnValAspPheAsnLysGluLeuLeuLeuLeu	216
Db	2061	AATGGAGTAAAAAGATGGACTAAAAAGTGGATTGTTTAAAAAGAGCTTCTCTGTGATT	2120
Qy	217	ProIleHisLeuGluValHisTrpSerLeuLeuSerValAspValArgArgThrIle	236
Db	2121	CCATATTCACCTGGAAGTCCACTGGTCTCTCATCTACTGTGACACTCTCTAATCGAATTATT	2180
Qy	237	ThrTyrPheAspSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyr	256
Db	2181	TCATTTTATGATTCACCAAGGCATTCAATTAAAGTTTTCGTAGAGATATAAAGAAAGTAT	2240
Qy	257	LeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyr	276
Db	2241	TTGCTGACTGAGCCAGAGAAAAAATAGCCTGAATTTCTTCAGGGTTGGCAGACTGCT	2300
Qy	277	PheLysMetAsnValAlaAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGln	296
Db	2301	GTTTACGAAGTGATTCCACACAGAAAAACACAGACTGACTGTGGAGTCTTTGTGCTCCAG	2360

```

Qy  297  TyTcysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLys  316
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  2361  TACTGCAAGTGCCTCGCCTTAGAGCAGCCTTTCCAGTTTTCACRAAGAAGACATGCCCGA  2420

Qy  317  LeuAargArgGlnIleTyrLysGlnLeuCysHisCysLysLeu  330
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
Db  2421  GTGCGGAAGAGGATTTACAAGGAGGTATGTGAGTGCCGCCTC  2462

Search completed: December 16, 2002, 22:23:25
Job time : 3338 secs

```

Search completed: December 16, 2002, 22:23:25
Job time : 3338 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2002, 19:39:15 ; Search time 46 Seconds
(without alignments)
2213.406 Million cell updates/sec

Title: US-09-848-852A-3

Perfect score: 1772

Sequence: 1 MYSQREWTGIIWARRGAHLA.....DMPKLRRIQYKELCHCKLTV 332

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPto1/US0984852/runat_16122002_132037_19035/app_query.fasta_1.519
-DB=Issued_Patents_NA -QMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0984852 -ACGN_1_1_38 @runat_16122002_132037_19035 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRRADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6 -FGAEXT=7
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	6.1	4835	4	Sequence 58, Appl
2	93	5.2	1983	1	Sequence 2, Appli
3	93	5.2	1983	1	Sequence 2, Appli
c	4	93	5.2	9840	Sequence 1, Appli
5	93	5.2	37895	1	Sequence 1, Appli
6	93	5.2	37895	1	Sequence 1, Appli
7	93	5.2	37895	3	Sequence 1, Appli
c	8	89.5	5.1	29793	Sequence 38, Appl
c	9	89.5	5.1	29793	Sequence 38, Appl
c	10	89.5	5.1	29793	Sequence 38, Appl
11	88.5	5.0	43676	3	Sequence 12, Appl
12	87.5	4.9	3456	4	Sequence 405, App

13	86	4.9	2184	4	US-09-134-001C-1230
14	84	4.7	925	2	US-08-531-927B-8
15	84	4.7	5342	3	US-08-902-632-1
16	84	4.7	5342	3	US-09-073-354-5
17	84	4.7	5342	3	US-08-656-005A-5
18	84	4.7	5342	3	US-09-073-259-5
19	84	4.7	5342	4	US-09-363-095-5
20	84	4.7	5342	4	US-09-418-027-5
21	83	4.7	903	1	US-08-469-665-1
22	83	4.7	903	2	US-09-038-596-1
23	83	4.7	903	5	PCT-US95-00421-1
24	83	4.7	1776	2	US-08-531-927B-1
25	83	4.7	1776	4	US-09-041-886-12
26	82.5	4.7	2023	3	US-08-961-083-199
27	82.5	4.7	32768	4	US-08-961-527-71
28	82	4.6	2900	1	US-08-034-650-9
29	82	4.6	2900	1	US-08-449-015-9
30	82	4.6	5339	3	US-09-073-354-6
31	82	4.6	5339	3	US-08-656-005A-6
32	82	4.6	5339	3	US-09-073-259-6
33	82	4.6	5339	4	US-09-363-095-6
34	82	4.6	5339	4	US-09-418-027-6
35	81.5	4.6	1902	3	US-09-041-991A-9
36	81	4.6	2007	4	US-09-134-001C-1979
37	81	4.6	11873	2	US-08-970-269A-32
38	81	4.6	11873	4	US-09-407-562-32
39	81	4.6	11878	2	US-08-970-269A-31
40	81	4.6	11878	4	US-09-407-562-31
41	81	4.6	11883	2	US-08-970-269A-28
42	81	4.6	11883	4	US-09-407-562-28
43	80.5	4.5	1350	4	US-09-134-001C-2836
44	80	4.5	873	2	US-08-531-927B-6
45	80	4.5	2395	4	US-08-961-527-205

ALIGNMENTS

RESULT 1

US-08-976-259-58
; Sequence 58, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:
LENGTH: 4835 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-976-259-58

Alignment Scores:

Pred. No.: 0.00929 Length: 4835
Score: 108.00 Matches: 59
Percent Similarity: 31.69% Conservative: 31
Best Local Similarity: 20.77% Mismatches: 86
Query Match: 6.09% Indels: 108
DB: 4 Gaps: 11

US-09-848-852A-3 (1-332) x US-08-976-259-58 (1-4835)

QY 42 LeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluProGlyGluLysAla 61
Db 1350 CTATTTCCCTCGCTCGCTAGGTATGTGACGATGTTCCGAACACCGACTTGGTGAAT 1409
QY 62 GlyClnHisSerProLeuArgGluGluHisVal-----ThrCysValGlnSerIle--- 78
Db 1410 CAAAGCCACAAATACCTGCATATGAACATATTTAGCTAGCTGCCAGAGACATAGAC 1469
QY 79 LeuAspGluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValVal 98
Db 1470 CTGCTGCATTTATAGTGTCTGATGAATTTTATGCCACTTGAATCGGAT----- 1520
QY 99 GluLysLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuVal 118
Db 1521 ACCAATTGTGCAGATTTTCTGAAAGACTTTCAGGATGAGAGTGCACGCTATTGCA 1580
QY 119 LeuClnLeuIleGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArg 138
Db 1581 TTG----- 1583
QY 139 ValAlaTyrlsArgHisValLeuThrMetAspLeuGlyThrLeuTyrglyGlnAsn 158
Db 1584 -----AAT 1586
QY 159 TrpLeuAsn----- 161
Db 1587 TGGCAAAATTTTGTCTAGTGTGAATGTTTCTGAAGAGGGTGGTTATTGAACGT 1646
QY 162 -----AspGlnValMetAsnMetTyrglyAspLeuValMetAspThrVal 176
Db 1647 TTTACCTATCGTCCCGCAATCTTTAAGCTTCATCATAACTTCAAAAGCGTGGTCAA 1706
QY 177 ProGluLysValHisPhePheAsnSerPhePheTyrlsAspLysLeuArgThrLysGlyTyr 196
Db 1707 CCGAGAGGAGTTAACCGCTTTCATATCCCATATGCTGATTTGGTATGCTGATAT 1766
QY 197 AspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuIle 216
Db 1767 -----ATCGATGCATTTGGTGTGCTGATTAATTTGATTCGTCAC 1799
QY 217 ProIleHis-----LeuGluValHisTrpSerLeuIleSerValAspVal 231
Db 1800 CCGAGGATGATTAATGGGGTGTAGTGTGAAGTACCTGGAGCGGTGTCAGGGTA----- 1853
QY 232 ArgArgArgThrIleThrTyrlsPheAspSerGlnArgThrLeuAsnArgArgCysProLys 251
Db 1854 -----AATCACTATGCGATTAA 1871
QY 252 His-----IleAlaLysTyrlsLeuGlnAlaGluAlaValLysLysAspArg 266
Db 1872 TCATTGAGGAATCTTGTGGCAAGCATCTGCGTGGTATGCTGCCACTGCTAATCGA 1931
QY 267 LeuAspPheHisGlnGlyTrpLysGlyTyrlsPheLysMetAsnValAlaArgGlnAsnAsn 286
Db 1932 GTAAAGCAT-----AAAGATTATTTCAAG-----GCACATGATCGTAAT 1970

QY 287 AspSerAspCys 290
Db 1971 GATGAAGAGTGC 1982

RESULT 2

US-08-375-709-2
Sequence 2, Application US/08375709

Patent No. 5683898

GENERAL INFORMATION:

APPLICANT: YAZAWA, Kazunaga

APPLICANT: YAMADA, Akiko

APPLICANT: KATO, Seishi

APPLICANT: KONDO, Kiyosi

TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid

TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of Eicosapentaenoic Acid

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/375,709

FILING DATE: 20-JAN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/178,251

FILING DATE: 14-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-147945

FILING DATE: 15-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: WEGNER, Harold C.

REGISTRATION NUMBER: 25,258

REFERENCE/DOCKET NUMBER: 53466/150/AAOK

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1983 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1983

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 1..1983

US-08-375-709-2

Alignment Scores:

Pred. No.: 0.169

Score: 93.00

Percent Similarity: 32.96%

Best Local Similarity: 20.95%

Query Match: 5.25%

DB: 1

Length: 1983

Matches: 75

Conservative: 43

Mismatches: 117

Indels: 123

Gaps: 19


```
US-09-848-852A-3 (1-332) x US-08-752-929-2 (1-1983)
Qy 9 GlyThrIleTrpAlaArgArgGlyAlaHisLeuAlaProProAspAlaSerIleLeuI 28
Db 537 GGACCACTTTGGCGGAGCTCGCGTGTCAAGAGATGTTCCCTGATGTCAAGACTACGG 596
Qy 28 eSerAsnValCysSerIleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAspLe 48
Db 597 CTCG-----GATACATCACTAAGAAATTTGTCGATGAGAACGTACT 638
Qy 48 uGly-----MetAlaGluAlaGluArgProGlyGlyAlaGlyGlnH1 64
Db 639 TGCGGTAACGCCATGAGCCGCGCAGCTTATCAATACGCGCAACACGTGGCAACA 698
Qy 64 s-----SerProLeuArgGluGluHisValThrCy 74
Db 699 TGACCAAGGTATTGTTGATGCTCGGTAGTAAAGTCTATCAAAAGGTGAATCACTTA 758
Qy 74 sValGlnSerIleLeuAspGluPheLeuGlnThrTyGlySerLeuIleProLeuSerTh 94
Db 759 CGTGCCTCA-----GACTACACCTTAACAGTGAAGCGAAATGGGAACGCTGAG-- 810
Qy 94 rAspGluValValGluLysLeuGluAspIlePheGlnGlnGlnPheSerThr----- 111
Db 811 -----ATTGATGGTCTAGAGATGGTGTATTATGGATGCTCGGCGCACCGAAGCTGA 860
Qy 112 -----ProSerArgLysGlyLeuValValLeuGlnLeuIleGlnSe 124
Db 861 GTCAGAAATGATCATTATATCCCTCTATAAAAGCGCTCTGGACGGCGAGCTT---AC 917
Qy 124 rTyGlnArgMetProGly----- 130
Db 918 CTATCAAGGTATCCACAACTTTATACGCTGCGCGCGCTAAAGTACGTGATCGGCTCAA 977
Qy 131 -----AsnAlaMetValArgGlyPhe-----ArgValAlaTy 141
Db 978 GTGGTCAAAAGATATCAACGAATATGATCAATGCTTTGGTCAAGATGTCGAAGTGTGT 1037
Qy 141 rLysArgHisVal----- 147
Db 1038 TGCTCGCACTTCGCGCAGTGTGGGTAAACGAGCATCAACGATTTCTTACGCGCTACA 1097
Qy 147 rMetAspLeuGlyThrLeuTyGlyGln-----AsnTrpLeuAsnAspGlnVal-- 164
Db 1098 GCGTGATACTACGCGCTAGTCACAACTTGAGACTTGCCAAAGCATGGTGTGCG 1157
Qy 165 -MetAsnMetTyGlyAspLeuValMetAspThrValProGluLysValHisPheAs 184
Db 1158 TATACAAGATATTGGCGATCGGATTCGAACACAGATTCAGAGTCTATC----- 1206
Qy 184 nSerPhePheTyAspLysLeuArgThrLysGlyTyAspGlyValLysArgTrpThrLy 204
Db 1207 -----TACAAGCGTGGCATACCAATGTTTACCACGCGACT-----TATAGCCA 1250
Qy 204 sAsnValAsp---IlePheAsnLysGluLeuLeuIleProIleHisLeuGluValH1 223
Db 1251 TAACGCTAAAGCGGTTTATACAAAGTATCTA----- 1281
Qy 223 sTrpSerLeuIleSerValAspValArgArgThrIleThrTyPheAspSerGlnAr 243
Db 1282 -----GGCTACTTCTGATATGAACCC 1301
Qy 243 gThrLeuAsnArgArgCysPro-----LysHisIleAlaLysTyLeuGln----- 258
Db 1302 AGCCAACCTTAATTCGCTGCCCAACCAAGCAAGTCTGCCAAGTTTGTCCGATACATGG 1361
Qy 259 -----AlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTy 276
Db 1362 CGCGGAGATGCGCAATTAAGCGGCTAAAGATGATTACGCTCAAGGTGAATACCGCTT 1421
Qy 276 rPhe-----LysMetAsnValAlaAlaArgGlnAsnAspSer 288
Db ----- 1421
```

```
Db 1422 TGTTCAAGCGCATTAATAAGGTGTGTGATGCGCGAGCCAGAAATCACTCC 1473
RESULT 4
US-09-534-638-1/c
; Sequence 1, Application US/09534638
; Patent No. 6320038
; GENERAL INFORMATION:
; APPLICANT: Panula, Pertti A.J.
; APPLICANT: Brandt, Annika
; APPLICANT: Westerlund, Johanna
; TITLE OF INVENTION: Promoter for Neuropeptide FF Promoter and use thereof
; TITLE OF INVENTION: for therapy and diagnosis
; FILE REFERENCE: 2530-104
; CURRENT APPLICATION NUMBER: US/09/534,638
; EARLIER FILING DATE: 2000-03-27
; EARLIER APPLICATION NUMBER: 09/365755
; EARLIER FILING DATE: 1999-08-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 9840
; TYPE: DNA
; ORGANISM: Mouse
; US-09-534-638-1
Alignment Scores:
Pred. No.: 2,12 Length: 9840
Score: 93.00 Matches: 48
Percent Similarity: 40.80% Conservative: 34
Best Local Similarity: 23.88% Mismatches: 67
Query Match: 5.25% Indels: 52
DB: 4 Gaps: 10
US-09-848-852A-3 (1-332) x US-09-534-638-1 (1-9840)
Qy 92 LeuSerThrAspGluValValGluLysLeuGluAspIlePhe----- 105
Db 3755 TTGGGAAAAGATGAGTTA-----AAACTGGAGGACATAGCCATCCTGACTGTGACAA 3702
Qy 106 -----GlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGln 120
Db 3701 AATTCACACAGAGAAACACAAAGAACATGAGACACAGACAGCGGCATCTA----- 3648
Qy 121 LeuIleGlnSerTyGlnArgMetPro-----GlyAsnAlaMetValArgGlyPheArg 138
Db 3647 -----CAACCTCTCCTTAATGAAAACCAAGAAAGTACTGAGAGTCTGAGA 3603
Qy 139 ValAlaTyLys-----ArgHisValLeuThrMetAspLeuGlyThrLeuTy 155
Db 3602 AAGAGTTCAGAGTTGATGTAGAGTCTCAGATACAAATGCATGAACTAGGAAGTGTATAC 3543
Qy 156 GlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyGlyAspLeuValMetAspThr 175
Db 3542 AAAACCTCCAAAATGAACAAAACAGTTG-----GTAAG 3510
Qy 176 ValProGluLysValHisPheAsnSerPhePheTyAspLysLeuArgThrLysGly 195
Db 3509 TTGGAGGACAAAGTCAGATATATATGGAGGAGATTTT----- 3474
Qy 196 TyrAspGlyValLysArgTrpThrLysAsn---ValAspIlePheAsnLysGlu----- 212
Db 3473 GATGAGGCGATTCAAACATGGAAAGGAACAAGATAGAAATTAACATGAAGAACAATCA 3414
Qy 213 -----LeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIle 227
Db 3413 AAGTGAACCCCTCACTAGAAAGTGTCCCAAGAGATTAAGATCATCGAGTTCT----- 3360
Qy 228 SerValAspValArgArgThrIleThrTyPheAspSerGlnArgThrLeuAsnArg 247
Db 3359 GCGGTGGACACAAAGTGGAAACAATCATATATCATGACACACACAGACAGACAGACA 3300
Qy 248 ArgCysProLysHisIleAlaLysTyLeuGlnAlaGluAlaValLysLysAspArgLeu 267
Db ----- 267
```


Db 3299 CAGACAGACACACACACACACACAGACAGACAGAGAGAGAGAGAGAGAGATA 3240

Qy 268 Asp 268

|||

Db 3239 GAT 3237

RESULT 5

US-08-375-709-1

; Sequence 1, Application US/08375709

; Patent No. 5683898

; GENERAL INFORMATION:

; APPLICANT: YAZAWA, Kazunaga

; APPLICANT: YAMADA, Akiko

; APPLICANT: KATO, Seishi

; APPLICANT: KONDO, Kiyosi

; TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid

; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of

; TITLE OF INVENTION: Eicosapentaenoic Acid

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/375,709

; FILING DATE: 20-JAN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/178,251

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 4-147945

; FILING DATE: 15-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: WEGNER, Harold C.

; REGISTRATION NUMBER: 25,258

; REFERENCE/DOCKET NUMBER: 53466/150/AAOK

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 37895 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:

; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM

; ORGANISM: BP-1625)

US-08-375-709-1

Alignment Scores:

Pred. No.: 17.8

Score: 37895

Matches: 75

Conservative: 43

Best Local Similarity: 20.96%

Mismatches: 117

Indels: 123

Gaps: 19

US-09-848-852a-3 (1-332) x US-08-375-709-1 (1-37895)

Qy 9 GlyThrIleTrp-AlaArgGlyAlaHisLeuAlaProProAspAlaSerIleLeuI1 28

||| :|||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

; Patent No. 5798259
; GENERAL INFORMATION:
; APPLICANT: YAZAWA, Kazunaga
; APPLICANT: YAMADA, Akiko
; APPLICANT: KATO, Seishi
; APPLICANT: KONDO, Kiyosi
; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing
; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic
; TITLE OF INVENTION: Acid
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,929
; FILING DATE: 20-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/375,709
; FILING DATE: 20-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/178,251
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-147945
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37895 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM
; ORGANISM: BP-1625)
; US-08-752-929-1

Alignment Scores:
Pred. No.: 17,8 Length: 37895
Score: 93.00 Matches: 75
Percent Similarity: 32.96% Conservative: 43
Best Local Similarity: 20.95% Mismatches: 117
Query Match: 5.25% Indels: 123
DB: 1 Gaps: 19

US-09-848-852a-3 (1-332) x US-08-752-929-1 (1-37895)

QY 9 GlyThrIleTrp-AlaAargArgGlyAlaHisLeuAlaProProAspAlaSerIleLeu1 28
DB 6657 GGACCACTTTGGCGGAGCTGCTCAAGAGATGTTCCCTCGATGTCAAAGTCTACGG 6716
QY 28 eSerAsnValCysSerIleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAsp1e 48
DB 6717 CTCA-----GATAACATCACTAAAGAAATTTGTCGATGAGAACGTACT 6758

QY 48 uGly-----MetAlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHi 64
DB 6759 TGCCGGTAACGCCATGAGCGCGCGCAGCTTATCAATAGCGGCAACACGTGCGCAACA 6818
QY 64 s-----SerProLeuArgGluGluHisValThrCy 74
DB 6819 TGACCACGTATTGTTGATGCTGCGCTAGTAAGTCTATCAAAAGGTGAATCACTTA 6878
QY 74 sValGlnSerIleLeuAspGluPheLeuGlnThrTyGlySerLeuIleProLeuSerTh 94
DB 6879 CGTCGCCCA-----GACTACACCTTAACAGCTGAAGCAATGGAAACCGTGCAG-- 6930
QY 94 rAspGluValValGluLysLeuAspIlePheGlnGlnGluPheSerThr----- 111
DB 6931 -----ATTGATGGTCTAGAGATGGTGTATTGATGGCTCGCGGACCGAAGCTGA 6980
QY 112 -----ProSerArgLysGlyLeuValLeuGlnLeuIleGlnSe 124
DB 6981 GTCAGAAATGATCATTATATCCCTCTAAAAAGCGCTCTGGACGGCGGAGCTT---AC 7037
QY 124 rTyGlnArgMetProGly----- 130
DB 7038 CTATCAAGGTATGCCACAACATTTATACGCTGCCGCGCGCTAAAGTAGCTGATCGGCTCAA 7097
QY 131 -----AsnAlaMetValArgGlyPhe-----ArgValAlaTy 141
DB 7098 GTGGTCAAAAGATATCAACGAAATGATCAATGCTTTGGTCAAGATGTCGAAGTGTGT 7157
QY 141 rLysArgHisVal-----LeuTh 147
DB 7158 TGCTCGCACTCTGCGCCAGTGTGGGTAACCAAGCGATCAACGATTTCTTACGCTACA 7217
QY 147 rMetAspAspLeuGlyThrLeuTyGln-----AsnTrpLeuAsnAspGlnVal-- 164
DB 7218 GCGTGATACTACGGCCTAGTGCACCAATCAACCTTGAGACTTGCACAGATGCTATC-- 7326
QY 184 nSerPhePheTyAspLysLeuArgThrLysGlyTyAspGlyValLysArgTrpThrLy 204
DB 7327 -----TACAAGACGTGTCATACCAATGTTTACCACCGCAGCT---TATAGCCA 7370
QY 204 sAsnValAsp---IlePheAsnLysGluLeuLeuIleProIleHisLeuGluValHi 223
DB 7371 TACGCTAAAGCGTTTATAACAAGTATCTA----- 7401
QY 223 sTrpSerLeuIleSerValAspValArgArgThrIleThrTyPheAspSerGlnAr 243
DB 7402 -----GGCTACTTCGATATGAACCC 7421
QY 243 gThrLeuAsnArgArgCysPro-----LysHisIleAlaLysTyLeuGln----- 258
DB 7422 AGCCAACTTAATCCGCTGCTCAACCAAGCAAGAAATCTGCCAAGTTTGTGGAATACATGGG 7481
QY 259 -----AlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTy 276
DB 7482 CGGCGCAGATGCGCAATTAAGCGCGCTAAAGATGATTCGCTCAAGGTGAATACCGCTT 7541
QY 276 rPhe-----LysMetAsnValAlaArgGlnAsnAsnAspSer 288
DB 7542 TGTTCGAACGGCATTAATAAGTGTGATGGCCGAGCCAGAAATGACTCC 7593

RESULT 7
US-09-090-793-1
; Sequence 1, Application US/09090793
; Patent No. 6140486
; GENERAL INFORMATION:
; APPLICANT: Calgene, LLC
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
; TITLE OF INVENTION: of polyketide-like synthesis genes in plants
; FILE REFERENCE: CGNE.131.01US

```

; CURRENT APPLICATION NUMBER: US/09/090,793
; CURRENT FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 37895
; TYPE: DNA
; ORGANISM: Shewanella putrefaciens
US-09-090-793-1

Alignment Scores:
Pred. No.: 17.8 Length: 37895
Score: 93.00 Matches: 75
Percent Similarity: 32.96% Conservative: 43
Best Local Similarity: 20.95% Mismatches: 117
Query Match: 5.25% Indels: 123
DB: 3 Gaps: 19

US-09-848-852A-3 (1-332) x US-09-090-793-1 (1-37895)

QY 9 GYThrIleTrp-AlaArGArgGLyAlaHisLeuAlaProProAspAlaSerIleLeuIl 28
||| :||| ||||| |||||
Db 6657 GGACCATTGGCGGAGCTCGCGGTTCACAGAGATGTCCTCATGAAGTCTACGG 6716

QY 28 eSerAsnValCysSerIleGlyAspHisValAlaGlnLeuPheGlnGlySerAspLe 48
||| ||| :||| :||| :||| :||| :||| :|||
Db 6717 CTCA-----GATAACATCACTAAAGAAATGTCGATGAGAACGTACT 6758

QY 48 uGly-----MetAlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHI 64
||| ||| :||| :||| :||| :||| :||| :|||
Db 6759 TGCCGGTAACGCCATGAGCGCGCGCAGCTATTCAATACGGCGCACACTGGGCAACA 6818

QY 64 s-----SerProLeuArgGluGluHisValThrCy 74
||| ||| :||| :||| :||| :||| :||| :|||
Db 6819 TGACCACGGTATTGTTGATGCTGCCTAGTAAAGTCTATCAAAAGGTGAATCACTTA 6878

QY 74 sValGlnSerIleLeuAspGluPheLeuGlnThrTyrglySerLeuIleProLeuSerTh 94
||| ||| :||| :||| :||| :||| :||| :|||
Db 6879 CGTCGCCCA-----GACTACACCTTAAACAGTCAAGGCAATGGGAACGCTGACG-- 6930

QY 94 rAspGluValValGluLysLeuGluAspIlePheGlnGlnPheSerThr----- 111
||||| :||| :||| :||| :||| :||| :|||
Db 6931 -----ATTGATGGTCTAGAGATGGTGTATTGATGCCCTCGGCACCCGAAGCTGA 6980

QY 112 -----ProSerArgLysGlyLeuValLeuGlnLeuIleGlnSe 124
||||| :||| :||| :||| :||| :||| :|||
Db 6981 GTCAGAAATGATCACTATATTCCTCTAAAAAAGCGCTCTGGACGCGGAGCTT---AC 7037

QY 124 rTyrglnArgMetProGly----- 130
||||| |||
Db 7038 CTATCAAGGTATGCCACAACATTTATACGCTCGCGCGCTAAAGTACGTGATGCGCTCAA 7097

QY 131 -----AsnAlaMetValArgGlyPhe-----ArgValAlaTy 141
||| ||| :||| :||| :||| :||| :|||
Db 7098 GTGGTCAAAAGATATCAACGAAATGATCAATGCCCTTTGGTCAAGATGTCGAAGTGTGT 7157

QY 141 rLysArgHisVal-----LeuTh 147
||||| |||
Db 7158 TGCTCGCATCTGGCCAGTGTGGGTAACCAACGATCAACGATTTCTTACGCCCTACA 7217

QY 147 rMetAspAspLeuGlyThrLeuTyrglyGln-----AsnTrpLeuAsnAspGlnVal-- 164
||||| :||| :||| :||| :||| :||| :|||
Db 7218 GCGTGATACTACGGCCTAGTCACAATCAACCTTGAGACTTCCCAACGATGTGTGTCGG 7277

QY 165 -MetAsnMetTyrglyAspLeuValMetAspThrValProGluLysValHisPheAs 184
||| ||| :||| :||| :||| :||| :||| :|||
Db 7278 TATACAAGATATTGGCGATGCGATTCAAGACACGATTCCAGACTATC----- 7326

QY 184 nSerPhePheTyrglyAspLysLeuArgThrLysGlyTyrglyValLysArgTrpThrLy 204
||||| ||| :||| :||| :||| :||| :||| :|||
Db 7327 -TACAGACGTGGCATCAATGAAATGGTTTACCAACGCTACT-----TATACCCA 7370

```

Db 26755 -----AAAATAATGAGGATCTTCAAAAAAATTTTATGATTGTGCCCATAT 26705
: : : : :
: : : : :
Qy 152 -----GlyThrLeuTyrGly-----GlnAsnTrpLeu 160
: : : : :
: : : : :
Db 26704 CCGGTTAACGGTTACGGTGAATATTAATTTGGTGATGAACACCGGAACAGAGAAGTTG 26645
: : : : :
: : : : :
Qy 161 AsnAspGlnValMetAsnMet-----TyrGlyAspLeu--- 171
: : : : :
: : : : :
Db 26644 ATGGATCAACTTTTGGCAAAATTAAGATTAATAATGATGGAGATTATTATGACATTTC 26585
: : : : :
: : : : :
Qy 172 ---ValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyrAsp--- 189
: : : : :
: : : : :
Db 26584 CACTTGCTGGAACATATTGACAAGATCATAGTACAGTTTCTCTTATTTCGAAGGATCCT 26525
: : : : :
: : : : :
Qy 190 -----LysLeuArgThrLysGlyTyrAspGly 198
: : : : :
: : : : :
Db 26524 ATACTTATACAAAAAGATGGAAGACAACACAGATCATTTCAAAACCTGGAAGACGAA 26465
: : : : :
: : : : :
Qy 199 ValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuPheProIle 218
: : : : :
: : : : :
Db 26464 GTTGACTATCTTACGAAAAAATTTGAAAGTTTGAATCCATCGCAGCTGGTTACATCTTG 26405
: : : : :
: : : : :
Qy 219 HisLeuGluValHisTrpSerLeuIleSerValAspVal---ArgArgArgThrIleThr 237
: : : : :
: : : : :
Db 26404 GAAACAAAGATGATTATCTCTTAGTGACAGTGGAAATTCCAAAAAAGAAAGATATTCG 26345
: : : : :
: : : : :
Qy 238 TyrPheAsp-SerGlnArgThrLeuAsnArg---ArgCysProLysHisIleAlaLysTy 256
: : : : :
: : : : :
Db 26344 TTTTCGAACTCAAGAAAAGATGAAGATTAATAATGCTTCAAGAATCCCTTCAATGTT 26285
: : : : :
: : : : :
Qy 256 rLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTy 276
: : : : :
: : : : :
Db 26284 TATCAGAAGGATAGTATTGAAAAG-----TCATGGAGAATTT 26245
: : : : :
: : : : :
Qy 276 rPheLysMetAsn 280
: : : : :
: : : : :
Db 26245 TTGGAACCTTCAAC 26233

RESULT 9
US-09-511-477-38/c
; Sequence 38, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; PRIOR FILING DATE: 2000-02-23
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 38
; LENGTH: 29793
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-477-38

Alignment Scores:
Pred. No.: 33.3 Length: 29793
Score: 89.50 Matches: 54
Percent Similarity: 41.51% Conservative: 56
Best Local Similarity: 20.38% Mismatches: 81
Query Match: 5.05% Indels: 74
DB: 4 Gaps: 13

US-09-848-852a-3 (1-332) x US-09-511-477-38 (1-29793)

Qy 59 GluLysAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIle 78
: : : : :
: : : : :
Db 26932 GAAAAAGATCAAAATATCAAAAAATTTAAAGAAGAA-----GTTGAGAAATTA 26885
: : : : :
: : : : :
Qy 79 LeuAspGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValVal 98
: : : : :
: : : : :
Db 26884 ATCGATGAA-----CAAGAAGATGAGATAATA 26858
: : : : :
: : : : :
Qy 99 -----GluLysLeuGluAspIlePheGlnGlnPheSerThrProSerArg----- 114
: : : : :
: : : : :
Db 26857 CTAGCTGAATGCAATTCGAAATTTTCCAAATAAAAGCAGACAGACAGCTAGAAATG 26798
: : : : :
: : : : :
Qy 115 LysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetVal 134
: : : : :
: : : : :
Db 26797 AGAAGCTTGTGATATGAGAAATGAGATTCTCAATCTC----- 26756
: : : : :
: : : : :
Qy 135 ArgGlyPheArgValAlaTyrLysArgHisValLeuThrMetAspAspLeu----- 151
: : : : :
: : : : :
Db 26755 -----AAAATAATGAGGATCTTCAAAAAAATTTTATGATTGTGCCCATAT 26705
: : : : :
: : : : :
Qy 152 -----GlyThrLeuTyrGly-----GlnAsnTrpLeu 160
: : : : :
: : : : :
Db 26704 CCGGTTAACGGTTACGGTGAATATTAATTTGGTGATGAACACCGGAACAGAGAAGTTG 26645
: : : : :
: : : : :
Qy 161 AsnAspGlnValMetAsnMet-----TyrGlyAspLeu--- 171
: : : : :
: : : : :
Db 26644 ATGGATCAACTTTTGGCAAAATTAAGATTAATAATGATGGAGATTATTATGACATTTC 26585
: : : : :
: : : : :
Qy 172 ---ValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyrAsp--- 189
: : : : :
: : : : :
Db 26584 CACTTGCTGGAACATATTGACAAGATCATAGTACAGCTTTTCTCTTATTTCGAAGGATCCT 26525
: : : : :
: : : : :
Qy 190 -----LysLeuArgThrLysGlyTyrAspGly 198
: : : : :
: : : : :
Db 26524 ATACTTATACAAAAAGATGGAAGACAACACAGATCATTTCAAAACCTGGAAGACGAA 26465
: : : : :
: : : : :
Qy 199 ValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuPheProIle 218
: : : : :
: : : : :
Db 26464 GTTGACTATCTTACGAAAAAATTTGAAAGTTTGAATCCATCGCAGCTGGTTACATCTTG 26405
: : : : :
: : : : :
Qy 219 HisLeuGluValHisTrpSerLeuIleSerValAspVal---ArgArgArgThrIleThr 237
: : : : :
: : : : :
Db 26404 GAAACAAAGATGATTATCTCTTAGTGACAGTGGAAATTCCAAAAAAGAAAGATATTCG 26345
: : : : :
: : : : :
Qy 238 TyrPheAsp-SerGlnArgThrLeuAsnArg---ArgCysProLysHisIleAlaLysTy 256
: : : : :
: : : : :
Db 26344 TTTTCGAACTCAAGAAAAGATGAAGATTAATAATGCTTCAAGAATCCCTTCAATGTT 26285
: : : : :
: : : : :
Qy 256 rLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTy 276
: : : : :
: : : : :
Db 26284 TATCAGAAGGATAGTATTGAAAAG-----TCATGGAGAATTT 26246
: : : : :
: : : : :
Qy 276 rPheLysMetAsn 280
: : : : :
: : : : :
Db 26245 TTGGAACCTTCAAC 26233

RESULT 10
US-09-511-507-38/c
; Sequence 38, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG)
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREAC
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 38
; LENGTH: 29793
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-507-38

Alignment Scores:
Pred. No.: 33.3 Length: 29793
Score: 89.50 Matches: 54
Percent Similarity: 41.51% Conservative: 56
Best Local Similarity: 20.38% Mismatches: 81
Query Match: 5.05% Indels: 74
DB: 4 Gaps: 13

US-09-848-852a-3 (1-332) x US-09-511-477-38 (1-29793)

QY	204	LysAsnValAspIlePheAsnLysGluLeuLeuLeuLeuProIleHisLeuGluValHis	223
DB	32796	GTTAACTGACCAACATTAAAGATATATTTTACGAACAACATTTTCATATA-----	32846
QY	224	TTPSerLeuIleSerValAspValArgArgThrIleThrTyPheAspSerGlnArg	243
DB	32847	TGGAAGCTA-----AGACATAAGAAGAAATTAACATCCACGGGTTCGANAAGA	32891
QY	244	ThrLeuAsnArgCysProLysHisIleAlaLysTyLeuGlnAlaGluAlaValLys	263
DB	32892	AGCATTTTCGAAGAATAAACACAGAAAGTTATTAACHTATGAATCAATAGAGCATTCGA	32951
QY	264	LysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyPheLysMetAsnValalaArg	283
DB	32952	GAAAAAGTGAGTGCCTTTTCTCTACAAAAGAAAATATCTGAATAAATGGGAAAAGAAAAC	33011
QY	284	GlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyPheLysHisLeu-----	301
DB	33012	ATTGAAAACGAAGATAAACTTTGGGSCACTTTGAACCTGGAGAATAAATTCATCAAACAA	33071
QY	302	-----AlaLeuSerGlnProPheSerPheThrGlnGlnAsp	313
DB	33072	RAGTTTTTCGCAAAATTAACCGGTCACTTTCAACATAGTCAACAAGAG	33119
RESULT 12			
US-09-134-001C-405			
; Sequence 405, Application US/09134001C			
; Patent No. 6380370			
; GENERAL INFORMATION:			
; APPLICANT: Lynn Doucette-Stamm et al			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAB			
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: GTC-007			
; CURRENT APPLICATION NUMBER: US/09/134,001C			
; CURRENT FILING DATE: 1998-08-13			
; PRIOR APPLICATION NUMBER: US 60/064,964			
; PRIOR FILING DATE: 1997-11-08			
; PRIOR APPLICATION NUMBER: US 60/055,779			
; PRIOR FILING DATE: 1997-08-14			
; NUMBER OF SEQ ID NOS: 5674			
; SEQ ID NO 405			
; LENGTH: 3456			
; TYPE: DNA			
; ORGANISM: Staphylococcus epidermidis			
US-09-134-001C-405			
Alignment Scores:			
Pred. No.: 1.97 Length: 3456			
Score: 87.50 Matches: 60			
Percent Similarity: 43.36% Conservative: 51			
Best Local Similarity: 23.44% Mismatches: 96			
Query Match: 4.94% Indels: 50			
DB: 4 Gaps: 14			
US-09-848-852A-3 (1-332) x US-09-134-001C-405 (1-3456)			
QY	98	ValGlnLysLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeu	117
DB	2581	ATAAATGAAATTAAGAAAAGTAGTATACAACTTTCCAAGGTTGAATCTAGAACAAATTTTA	2640
QY	118	ValLeuGlnLeuIleGlnSerTyPheGlnArgMetProGlyAsn-----AlaMetValArg	135
DB	2641	GAGCAGCAGATGACTTTATCAATTCGACAAATGAATAACTCTTTAAAAAATGATTTAA	2700
QY	136	GlyPhe-----ArgValAlaTyPheLysArgHisValLeuThrMetAspAsp	150
DB	2701	GATTTTTCGATGACAAACGCCAAATTTATCAGCTAGACAACAT-----AAA	2745
QY	151	LeuGlyThrLeuTyGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyPheGlyAsp	170
DB	2746	TTAAATCACATATAAAAACCACCAACCCCTATTAAATCAAGAGTTGATTTGATGACAACTGCACA	2805
QY	171	LeuValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyPheAspLys	190

Db	2806	CCTACTTTCAATGAAGTAGAAGAACAGTATATCTCTAAAT		: :		: :	GAACGG	2853
Qy	191	LeuArgThrLysGlyTyrAspGlyVallys	-----	ArgTrpThrLysAsn	205	: :		
Db	2854	TTAAACATACAACTTTTACGTAGAGTTAAATCTGTGTTAATAGTCAGATGACACAAAT	2913	: :				
Qy	206	ValAspIlePheAsnLysGlu	-----	LeuLeuLeuIlePheIleHis	219	: :		
Db	2914	AACGAC	---TTAATGAGCAAAAGAAAATTTCACTAAATATATTTAGATCAAAATTCAT	2970	: :			
Qy	220	-----	LeuGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThr	237	: :			
Db	2971	CAACGCTTTATCTTAGAGCAATCACTTATCACAGAAAGATTAAAAA	3018	: :				
Qy	238	TyrPheAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeu	257	: :				
Db	3019	TATTTTAAATTCACAA	-----CTAGAAGAACAAATCTCACCAGTCATCAAAAAGTTAAAT	3072	: :			
Qy	258	GlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGly	Trp	-----LysGln	275	: :		
Db	3073	CAGATTCAATGTTATTAATAATGCAAAATTTAATGTTGAGCCATCACTGGTGTGATACGGCG	3132	: :				
Qy	275	yTyrPheLysMetAsnValAlaArg	-----GlnAsnAsnAspSerAspCys	290	: :			
Db	3133	TTACTTCAAAATGAACTTAAATTCAAATGTTGCAATCACTACCAAAACAGTTAACTAAACGT	3192	: :				
Qy	290	sGlyAlaPheValLeuGlnTyrCysLys	-----HisLeuAlaLeuSerGlnPr	306	: :			
Db	3193	AAATAGTAAATCCAAAGTCACAA	---AAGGATATACAAAGTACACATAGCT	---AATCAAAC	3248	: :		
Qy	306	oPheSerPheThrGlnGlnAspMetProLysLeuArgGlnIle	321	: :				
Db	3249	TCTTGAATTATACAAAGTAGATTGAACTCATTCGCGGACAAATTA	3294	: :				
RESULT 13								
US-09-134-001C-1230								
; Sequence 1230, Application US/09134001C								
; Patent No. 6380370								
; GENERAL INFORMATION:								
; APPLICANT: Lynn Doucette-Stamm et al								
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS								
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS								
; FILE REFERENCE: GPC-007								
; CURRENT APPLICATION NUMBER: US/09/134,001C								
; CURRENT FILING DATE: 1998-08-13								
; PRIOR APPLICATION NUMBER: US 60/064,964								
; PRIOR FILING DATE: 1997-11-08								
; PRIOR APPLICATION NUMBER: US 60/055,779								
; PRIOR FILING DATE: 1997-08-14								
; NUMBER OF SEQ ID NOS: 5674								
; SEQ ID NO 1230								
; LENGTH: 2184								
; TYPE: DNA								
; ORGANISM: Staphylococcus epidermidis								
US-09-134-001C-1230								
Alignment Scores:								
Pred. No.: 1.47 Length: 2184								
Score: 86.00 Matches: 49								
Percent Similarity: 36.26% Conservative: 46								
Best Local Similarity: 18.70% Mismatches: 93								
Query Match: 4.85% Indels: 74								
DB: 4 Caps: 9								
US-09-848-852A-3 (1-332) x US-09-134-001C-1230 (1-2184)								
Qy	69	GluGluHisValThr	---CysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGly	87	: :	: :		
Db	58	GAGGAATACATCACAGTGTCTAGAGAGTATTAAAGCAACACGCTACACAGATTTAAT	117	: :	: :	: :		
Qy	88	SerLeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGln	107	: :				

Search completed: December 16, 2002, 20:17:51
Job time : 115 secs

1

1

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.
OM protein - protein search, using sw model
Run on: December 16, 2002, 19:34:35 ; Search time 771 Seconds
(without alignments)
88.726 Million cell updates/sec

Title: US-09-848-852A-3
Perfect score: 1772
Sequence: 1 MYSQRFWGTIWARAGHA.....DMPKLRROIYKELCHCKLTV 332

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archheap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1667	94.1	371	Q9Y3W9	Q9Y3W9 homo sapien
2	1667	94.1	574	Q9H4L4	Q9H4L4 homo sapien
3	1667	94.1	574	Q96PS4	Q96PS4 homo sapien
4	1637	92.4	568	Q9EP97	Q9EP97 mus musculus
5	789	44.5	537	Q96H10	Q96H10 homo sapien
6	788.5	44.5	755	Q96P32	Q96P32 macaca fasc
7	784	44.2	446	Q96SA5	Q96SA5 homo sapien
8	490	27.7	191	Q9NWF3	Q9NWF3 homo sapien
9	456	25.7	643	Q9P003	Q9P003 homo sapien
10	393	22.2	412	Q9P2L5	Q9P2L5 homo sapien
11	393	22.2	589	Q96SR2	Q96SR2 homo sapien
12	393	22.2	590	Q9HC62	Q9HC62 homo sapien
13	387.5	21.9	357	Q9CUM5	Q9CUM5 mus musculus
14	385.5	21.8	541	Q91ZX6	Q91ZX6 mus musculus
15	385.5	21.8	588	Q9D4Z0	Q9D4Z0 mus musculus
16	380	21.4	588	Q9EQE1	Q9EQE1 rattus norv

17	311	17.6	468	5	Q9VP10	Q9VP10 drosophila
18	307	17.3	1513	5	Q9VWK5	Q9VWK5 drosophila
19	282.5	15.9	497	10	Q9FP66	Q9FP66 oryza sativ
20	266.5	15.0	155	11	Q9WV12	Q9WV12 mus musculus
21	260.5	14.7	478	10	Q9M908	Q9M908 arabidopsis
22	260.5	14.7	489	10	Q94F30	Q94F30 arabidopsis
23	253.5	14.3	100	6	Q95KG4	Q95KG4 macaca fasc
24	253.5	14.3	244	5	Q8SVK1	Q8SVK1 encephalito
25	252	14.2	157	5	O01650	O01650 drosophila
26	240.5	13.6	233	10	O65278	O65278 arabidopsis
27	201.5	11.4	571	10	Q8RWNO	Q8RWNO arabidopsis
28	201.5	11.4	582	10	Q9X1J4	Q9X1J4 arabidopsis
29	197.5	11.1	424	10	O23439	O23439 arabidopsis
30	193	10.9	547	10	O80745	O80745 arabidopsis
31	180	10.2	506	10	O8S2J1	O8S2J1 oryza sativ
32	173.5	9.8	1017	4	Q96PS5	Q96PS5 homo sapien
33	169.5	9.6	1148	5	Q9VRY4	Q9VRY4 drosophila
34	163	9.2	710	10	O81879	O81879 arabidopsis
35	156	8.8	1240	3	Q9P6U5	Q9P6U5 neurospora
36	154	8.7	560	5	Q9VYJ5	Q9VYJ5 drosophila
37	143.5	8.1	1011	10	O8S1I4	O8S1I4 oryza sativ
38	143	8.1	1105	4	O8TBY4	O8TBY4 homo sapien
39	137.5	7.8	997	10	Q8RZ25	Q8RZ25 oryza sativ
40	137.5	7.8	1007	10	Q94LG5	Q94LG5 oryza sativ
41	136.5	7.7	1046	10	O8W0H2	O8W0H2 oryza sativ
42	136	7.7	893	5	Q23238	Q23238 caenorhabdi
43	135.5	7.6	1053	10	O94E62	O94E62 oryza sativ
44	135	7.6	226	10	Q9LSS7	Q9LSS7 arabidopsis
45	135	7.6	1011	10	Q8W5N9	Q8W5N9 oryza sativ

ALIGNMENTS

RESULT 1

Q9Y3W9 PRELIMINARY; PRT; 371 AA.
AC Q9Y3W9;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 42.2 kDa protein.
GN DKFZP586K0919.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Cassenhuber J., Glassl S.,
RA Ansgorge W., Boecker M., Bloecher H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
RA Mewes H.W., Othenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL Genome Res. 11:422-435(2001).
DR EMBL; AL050283; CAB43384.1; -.
DR HSSP; Q02724; 1EUV.
DR MEROPS; C48.003; -.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Hypothetical protein.
SQ SEQUENCE 371 AA; 42234 MW; EEB87C5885CA5082 CRC64;
Query Match 94.1%; Score 1667; DB 4; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.7e-136; Indels 0;
Matches 314; Conservative 0; Mismatches 0; Gaps 0;
QY 19 LAPPDASILISVCSIGDHVAQELFGQSDLGMAEAEAEKAGQHSPLREHVTVCQSI 78
|||||

```
Db 58 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 117
QY 79 LDEFLOTYGLSLPLSTDEVVEKLEDFIQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 138
Db 118 LDEFLOTYGLSLPLSTDEVVEKLEDFIQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 177
QY 139 VAYKRHLVTMDLGLTLYGQNLNDQVNMVYGDVMTVPKVFHFFNSFFYDKLRTKGYDG 198
Db 178 VAYKRHLVTMDLGLTLYGQNLNDQVNMVYGDVMTVPKVFHFFNSFFYDKLRTKGYDG 237
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258
Db 238 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 297
QY 259 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLYQCKHLALSOPFSFTQODMPKLR 318
Db 298 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLYQCKHLALSOPFSFTQODMPKLR 357
QY 319 RQIYKELCHCKLTV 332
Db 358 RQIYKELCHCKLTV 371

RESULT 2
Q9H4L4
ID Q9H4L4 PRELIMINARY; PRT; 574 AA.
AC Q9H4L4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sentrin/SUMO-specific protease.
GN SBNP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yeh E.T., Gong L., Kamitani T.;
RA "Ubiquitin-like proteins: new wines in new bottles.";
RL Gene 248:1-14(2000).
DR EMBL; AY008763; AAG33252.1; -.
DR HSSP; Q02724; LEUV.
DR MEROPS; C48.003; -.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 574 AA; 64856 MW; 830FD39C4D02C0EB CRC64;

Query Match 94.1%; Score 1667; DB 4; Length 574;
Best Local Similarity 100.0%; Pred. No. 3.2e-136;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 78
Db 261 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 320
QY 79 LDEFLOTYGLSLPLSTDEVVEKLEDFIQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 138
Db 321 LDEFLOTYGLSLPLSTDEVVEKLEDFIQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 380
QY 139 VAYKRHLVTMDLGLTLYGQNLNDQVNMVYGDVMTVPKVFHFFNSFFYDKLRTKGYDG 198
Db 381 VAYKRHLVTMDLGLTLYGQNLNDQVNMVYGDVMTVPKVFHFFNSFFYDKLRTKGYDG 440
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258
Db 441 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 500
QY 259 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLYQCKHLALSOPFSFTQODMPKLR 318
Db 501 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLYQCKHLALSOPFSFTQODMPKLR 560

RESULT 3
Q96PS4
ID Q96PS4 PRELIMINARY; PRT; 574 AA.
AC Q96PS4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SUMO-1 specific protease 3.
GN SSP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi S.J., Jeon Y.J., Kim K.I., Nishimori S., Suzuki T., Uchida S.,
RA Shimbara N., Tanaka K., Chung C.H.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199459; AAL25652.1; -.
DR MEROPS; C48.003; -.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 574 AA; 65009 MW; E495137EE7500741 CRC64;

Query Match 94.1%; Score 1667; DB 4; Length 574;
Best Local Similarity 100.0%; Pred. No. 3.2e-136;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 78
Db 261 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 320
QY 79 LDEFLOTYGLSLPLSTDEVVEKLEDFIQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 138
Db 321 LDEFLOTYGLSLPLSTDEVVEKLEDFIQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 380
QY 139 VAYKRHLVTMDLGLTLYGQNLNDQVNMVYGDVMTVPKVFHFFNSFFYDKLRTKGYDG 198
Db 381 VAYKRHLVTMDLGLTLYGQNLNDQVNMVYGDVMTVPKVFHFFNSFFYDKLRTKGYDG 440
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258
Db 441 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 500
QY 259 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLYQCKHLALSOPFSFTQODMPKLR 318
Db 501 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLYQCKHLALSOPFSFTQODMPKLR 560

RESULT 4
Q9EP97
ID Q9EP97 PRELIMINARY; PRT; 568 AA.
AC Q9EP97;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sentrin/SUMO-specific protease (SMT3 isopeptidase 1).
GN SMT3IP1 OR SEMP3 OR SMT3IP.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
```

```
QY 319 RQIYKELCHCKLTV 332
Db 561 RQIYKELCHCKLTV 574

RESULT 3
Q96PS4
ID Q96PS4 PRELIMINARY; PRT; 574 AA.
AC Q96PS4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SUMO-1 specific protease 3.
GN SSP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi S.J., Jeon Y.J., Kim K.I., Nishimori S., Suzuki T., Uchida S.,
RA Shimbara N., Tanaka K., Chung C.H.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199459; AAL25652.1; -.
DR MEROPS; C48.003; -.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 574 AA; 65009 MW; E495137EE7500741 CRC64;

Query Match 94.1%; Score 1667; DB 4; Length 574;
Best Local Similarity 100.0%; Pred. No. 3.2e-136;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 78
Db 261 LAPPDASILSNVCSIGDHVAQELFGQSDLGMAEAEAPGEGKAGQHSPLREEHVTCVQSI 320
QY 79 LDEFLOTYGLSLPLSTDEVVEKLEDFIQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 138
Db 321 LDEFLOTYGLSLPLSTDEVVEKLEDFIQEFSTPSRKGVLVLQIQTQYQRMGPNAMVRGFR 380
QY 139 VAYKRHLVTMDLGLTLYGQNLNDQVNMVYGDVMTVPKVFHFFNSFFYDKLRTKGYDG 198
Db 381 VAYKRHLVTMDLGLTLYGQNLNDQVNMVYGDVMTVPKVFHFFNSFFYDKLRTKGYDG 440
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 258
Db 441 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQ 500
QY 259 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLYQCKHLALSOPFSFTQODMPKLR 318
Db 501 AEAVKKDRDLDFHOGWKGYFKMNVARQNNSDCGAFVLYQCKHLALSOPFSFTQODMPKLR 560

RESULT 4
Q9EP97
ID Q9EP97 PRELIMINARY; PRT; 568 AA.
AC Q9EP97;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sentrin/SUMO-specific protease (SMT3 isopeptidase 1).
GN SMT3IP1 OR SEMP3 OR SMT3IP.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
```

```

[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=20267842; PubMed=10806345;
RA Yeh E.T., Gong L., Kamitani T.;
RT "Ubiquitin-like proteins: new wines in new bottles.";
RL Gene 248:1-14(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed=11029585;
RA Nishida T., Tanaka H., Yasuda H.;
RT "A novel mammalian Smt3-specific isopeptidase 1 (SMT3IP1) localized in
the nucleolus at interphase.";
RL Eur. J. Biochem. 267:6423-6427(2000).
DR EMBL; AY008764; AAG33253.1; -;
DR EMBL; AF194031; AAG28418.1; -;
DR HSSP; Q02724; IEUV.
DR MEROPS; C48.003; -;
DR MGD; MGI:2158736; Smt3ip1.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 568 AA; 64403 MW; 655F1FABIAB62EA8 CRC64;

Query Match 92.4%; Score 1637; DB 11; Length 568;
Best Local Similarity 98.1%; Pred. No. 1.3e-133;
Matches 308; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 19 LAPPDASILISNVCISGDHVAQELFGQSDLGMAEAEAPRGKAGQHSPLREEHVTCVQSI 78
DB 255 LAPPDASILISNVCISGDHVAQELFGQSDLGIAEADRTGKAGQHSPLREEHVTCVQSI 314
QY 79 LDEFLOTGSLPLSTDEVEKLEDIFQOEFSTPSRKGLVLQLIQSYORMPGNMVRGR 138
DB 315 LDEFLOTGSLPLSTDEVEKLEDIFQOEFSTPSRKGLVLQLIQSYORMPGNMVRGR 374
QY 139 VAYKRHLVTMDLGTLYGQNLNDQVMNMYGDLVMDTVPEKVFHFFNSFYDKLRKGYDG 198
DB 375 VSYKRHLVTMDLGTLYGQNLNDQVMNMYGDLVMDTVPEKVFHFFNSFYDKLRKGYDG 434
QY 199 VKRWTKNVDIFNKKELLIPILHLEVHWSLISVDVRRRTTYFDQSRTLNRCPKHIKYLQ 258
DB 435 VKRWTKNVDIFNKKELLIPILHLEVHWSLISVDVRRRTTYFDQSRTLNRCPKHIKYLQ 494
QY 259 ABAAVKDRLDHFQGWKGYFKMNVARQNDSDCGAFVLYQCKHLALSQPFSFTQDDMPKLR 318
DB 495 ABAAVKDRLDHFQGWKGYFKMNVARQNDSDCGAFVLYQCKHLALSQPFSFTQDDMPKLR 554
QY 319 ROIYKELCHCKLTIV 332
DB 555 ROIYKELCHCKLTIV 568

RESULT 5
Q96HI0 PRELIMINARY; PRT; 537 AA.
AC Q96HI0;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Unknown (Protein for IMAGE:3448367) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008589; AA080589.1; -;
DR MEROPS; C48.008; -;
DR InterPro; IPR003653; SUMO_protease.
```

```

DR Pfam; PF02902; Peptidase_C48; 1.
FT NON_TER 1
SQ SEQUENCE 537 AA; 61769 MW; 18F609C41D4B7DF9 CRC64;

Query Match 44.5%; Score 789; DB 4; Length 537;
Best Local Similarity 54.0%; Pred. No. 5.1e-60;
Matches 148; Conservative 54; Mismatches 66; Indels 6; Gaps 3;

QY 58 GEKAGQHSPLREHVT-CVQSTLDFLOTYGSILPLSTDEVEKLEDIFQOEFSTPSRK 116
DB 267 GNSKASQSPVDDEQSVCLSGFLDEVMKYGSLVPLSEKEVLGRKLDVFNEDFS--NRKP 324
QY 117 LVLQLIQSYQRMPPGNMVRGRVAYKRVHVTMDLGTLYGQNLNDQVMNMYGDLVMDTV 176
DB 325 FINREITNYRARKQC---NFRIFYNKKHMLDMDLATLDGQNLNDQVINMYGELIMDAV 381
QY 177 PEKVHFFNSFYDKLRKGYDGKRWTKNVDIFNKKELLIPILHLEVHWSLISVDVRRRTI 236
DB 382 PDKVHFFNSFYHRLVTKGYNGVKRWTKVDLFRKSLLLIPILHLEVHWSLITVITLSNRII 441
QY 237 TVFDSQRTLNRRCPKHIKYLQAEAVKVDRLDFHOGWKGYFKMNVARQNDSDCGAFVLYQ 296
DB 442 SFYDSQGIHFKFCVENIRKYLTEAREKNRPEFLQGWQVAVTKCIPQQKNDSDCGVFLVQ 501
QY 297 YCKHLALSQPFSFTQDDMPKLRRIYKELCHCKL 330
DB 502 YCKCLALEQPFQSOEDMPRVKRIYKELCECRL 535

RESULT 6
Q8WP32 PRELIMINARY; PRT; 755 AA.
AC Q8WP32;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical 86.3 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_Taxid=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
libraries.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB074445; BAB72076.1; -;
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Hypothetical protein.
SQ SEQUENCE 755 AA; 86290 MW; 40EC773CA29B8CEA CRC64;

Query Match 44.5%; Score 788.5; DB 6; Length 755;
Best Local Similarity 53.5%; Pred. No. 9.2e-60;
Matches 152; Conservative 53; Mismatches 72; Indels 7; Gaps 4;

QY 49 GMAEEAERPERKAGQH-SPLREHVT-CVQSTLDFLOTYGSILPLSTDEVEKLEDIFQ 106
DB 475 GLKLENQVGGKDSQKASFPVDEQLSVCLSGFLDEVMKYGSLVPLSEKEVLGRKLDVFN 534
QY 107 QEFSTPSRKGLVLQLIQSYQRMPPGNMVRGRVAYKRVHVTMDLGTLYGQNLNDQVMN 166
DB 535 EDFS--NRKPFINREITNYRARKQC---NFRIFYNKKHMLDMDLATLDGQNLNDQVIN 589
QY 167 MYGDLVMDTVPEKVHFFNSFYDKLRKGYDGKRWTKNVDIFNKKELLIPILHLEVHWSL 226
DB 590 MYGELIMDAVPDKVHFFNSFYHRLVTKGYNGVKRWTKVDLFRKSLLLIPILHLEVHWSL 649
QY 227 ISVDVRRRTIITYFDSQRTLNRRCPKHIKYLQAEAVKVDRLDFHOGWKGYFKMNVARQNN 286
```

```

Db 650 ITVTLSNRIISFYDSQIHFKFCVENIRKYLLEAREKNRPEFLQGWQTAVTKCIPOOKN 709
QY 287 DSDCGAVLOYCKHALSQSFSTQDMPKLRQIYKELCHCKL 330
Db 710 DSDCGVEVLOYCKLALEQPFQSFQSDMPRVRKRIYKELCECL 753

RESULT 7
Q96SA5
ID Q96SA5 PRELIMINARY; PRT; 446 AA.
AC Q96SA5;
DT 01-OCT-2001 (TrEMBLrel. 19, Created)
DT 01-DSC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sumo/sentrin-specific protease.
GN FKSG45.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang Y.-G., Li T.;
RT "Identification of FKSG45, a novel gene located on human chromosome
RT 3."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF335474; AK69630.1;
DR MROPS; C48.008;
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 446 AA; 50902 MW; 3810DCC23655D646 CRC64;

Query Match 44.2%; Score 784; DB 4; Length 446;
Best Local Similarity 53.6%; Pred. No. 1.1e-59;
Matches 147; Conservative 54; Mismatches 67; Indels 6; Gaps 3;

QY 58 GERAGQHSPLREHVT-CVQSIIDFLQTYGSLIPSTDEVEKLEDFQOEFSTPSRK 116
Db 176 GKNSQKASPVDDQSLVCLSGFLDEWKKYKGLSLVSEKEVLGRKLDVFNEDFC--NRKP 233
QY 117 LVQLQISQYQRMFGNMRVGRVAYKRHLVLTMDLGLTYGNLNDQVMNMGDLVMDTV 176
Db 234 FINREITNRYARKQC--NFRIFYNKHMLDMDLDTLQGNLNDQVINNYGELIMDAV 290
QY 177 PEKVHFNFSFYDKLRTKGVGKVRKTKNVDFENKELLIPILHLEVHWSLISVDVRRRTI 236
Db 291 PDKVHFNFSFRLQVTKGYNGYKRWTKKYDLEFKKSLILTIPIHLEVHWSLITVLSNRRI 350
QY 237 TYFDSORTLNRCPKHTAKYLOAEAVKDRDLDFHOGKKGFKMNVARQNNDSCGAFVLQ 296
Db 351 SFYDSQGIHFKFCVENIRKYLLEAREKNRPEFLQGWQTAVTKCIPOOKNDSGCVFVLQ 410
QY YCKHLALSQSFSTQDMPKLRQIYKELCHCKL 330
Db 411 YCKCLALEQPFQSFQSDMPRVRKRIYKELCECL 444

RESULT 8
Q9NWF3
ID Q9NWF3 PRELIMINARY; PRT; 191 AA.
AC Q9NWF3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Hypothetical 20.5 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RC TISSUE=EMBRYO;
RA Isoai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK00923; BAA91428.1;
SQ SEQUENCE 191 AA; 20457 MW; B9F2B6F629F31612 CRC64;

Query Match 27.7%; Score 490; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.1e-34;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILISNVCISGDHVAQELFOGSDLGMAEAEERPGEKAGQHSPLREHVTQVSI 78
Db 96 LAPPDASILISNVCISGDHVAQELFOGSDLGMAEAEERPGEKAGQHSPLREHVTQVSI 155
QY 79 LDEFLOTYGSLLPLSTDEVVEKLEDFQOEFSTPSR 114
Db 156 LDEFLOTYGSLLPLSTDEVVEKLEDFQOEFSTPSR 191

RESULT 9
Q9P0U3
ID Q9P0U3 PRELIMINARY; PRT; 643 AA.
AC Q9P0U3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sentrin/SUMO-specific protease.
GN SENP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20119292; PubMed=10652325;
RA Gong L., Millas S., Maul G.G., Yeh E.T.H.;
RT "Differential Regulation of Sentrinized Proteins by a Novel Sentrin-
RT specific Protease."
RL J. Biol. Chem. 275:3355-3359(2000).
DR EMBL: AF149770; AAF31171.1;
DR HSSP; Q02724; LEUV.
DR MEROPS; C48.002;
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
KW Protease.
SQ SEQUENCE 643 AA; 73351 MW; 1556864F1BDEE337 CRC64;

Query Match 25.7%; Score 456; DB 4; Length 643;
Best Local Similarity 34.0%; Pred. No. 5.5e-31;
Matches 112; Conservative 60; Mismatches 111; Indels 46; Gaps 11;

QY 24 ASLISNVCISGDHVAQE----LFQSDLGMAEAEERPGEKAGQHS-----PLR 68
Db 338 AELWIKELTSYDSRARELRQIEQKALQALQNLQER--EHSVHDSVELHLRVPLE 395
QY 69 EE-HVTQVOSILDEFLOTYGSLLPLSTDEVVEKLEDFQOEFSTPSRKGLVLQLQSYOR 127
Db 396 KEIPVTIVQVE-----TQKKGKHLTDEDEFEPEITEM-EKEIKNVFRNG----- 438
QY 128 MFGNAMVRGFRVAYKRHLVLTMDLGLTYGNLNDQVMNMGDLVMDTVPEK----VHFF 183
Db 439 NQDEVLSAEFL-----TITRKDIQTLNHLNWLNDIINFYNNMLMERKGLPSVHAF 493
QY 184 NSFFYDKLTKTGVDGKVRKTKNVDFENKELLIPILHLEVHWSLISVDVRRRTTYDSOR 243
Db 494 NTFFFTKLKTAGYQAVKRWTKKYDVFESVDILLVPIHLGVHCLAVVDYFRKKNITYDSMG 553
```



```
Db 319 -EFTDMEKEI-----SNAHGPGPPDEILSSAFKLRI-TRGDIQTLKNYHNLN 364
QY 162 DOVMNMYGDLVMDTVPEK----VHFNSFFYDKLRTKGYDGVKRWTKNVDIENKELLIP 217
Db 412 DEVINFYNLLVERSKKQGYPALHAFSTFFYPKLKSGGYQAVKRWTKGVNLEFEQELVLP 471
QY 218 IHLEVHWSLISVDVRRRTITYPDSQRTLNRRCPKHIKYLQAEAVKKRDLDFH-OGWKGY 276
Db 472 IHRKVHWSLVVMDLRKKCLKYLDSMGQKGRHICEILLQYLQDESCKTKRNTDNLLEWTHY 531
QY 277 FKM--NVARQNNDSDCGAFVLOYCKHLALSOPFSTQODMPKLRRQIYKELCHCKL 330
Db 425 IHRKVHWSLVVMDLRKKCLKYLDSMGQKGRHICEILLQYLQDESCKTKRNTDNLLEWTHY 484
QY 277 FKM--NVARQNNDSDCGAFVLOYCKHLALSOPFSTQODMPKLRRQIYKELCHCKL 330
Db 485 SMKPEIPQOLNGSDCGMFTCKYADYISRDKITFTQHOMPLFRKKMWWEILHQOL 540

RESULT 15
Q9D4Z0
ID Q9D4Z0 PRELIMINARY; PRT; 588 AA.
AC Q9D4Z0;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE 4930538C18Rik protein.
GN 4930538C18RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischer W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., King B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015987; BAB30067.1; -.
DR HSSP; Q02724; 1EUV.
DR MEROPS; C48.007; -.
DR MGD; MGI:1923076; 4930538C18Rik.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
SQ SEQUENCE 588 AA; 67579 MW; 09B56796CA194847 CRC64;

Query Match 21.8%; Score 385.5; DB 11; Length 588;
Best Local Similarity 31.4%; Pred. No. 6.4e-25;
Matches 93; Conservative 62; Mismatches 106; Indels 35; Gaps 9;

QY 55 ERPGKAGQHSPLREHVTCT--VQSILDEFQTGSLIPLS-----TDEVVEKLEDFQ 106
Db 307 EREGTRGHQMEPDLSEVSARLGLSGSNGLLRRKISLVLEIKEKNFPSPKEDRPTEDLF- 365
QY 107 QEFSTPSRKLVLQLIQSVQRMGNMVRG-----FRVAYKRHYLTMDDLGTLGYQNWLN 161
Db 366 -EFTDMEKEI-----SNAHGPGPPDEILSSAFKLRI-TRGDIQTLKNYHNLN 411
```

```
QY 162 DOVMNMYGDLVMDTVPEK----VHFNSFFYDKLRTKGYDGVKRWTKNVDIENKELLIP 217
Db 412 DEVINFYNLLVERSKKQGYPALHAFSTFFYPKLKSGGYQAVKRWTKGVNLEFEQELVLP 471
QY 218 IHLEVHWSLISVDVRRRTITYPDSQRTLNRRCPKHIKYLQAEAVKKRDLDFH-OGWKGY 276
Db 472 IHRKVHWSLVVMDLRKKCLKYLDSMGQKGRHICEILLQYLQDESCKTKRNTDNLLEWTHY 531
QY 277 FKM--NVARQNNDSDCGAFVLOYCKHLALSOPFSTQODMPKLRRQIYKELCHCKL 330
Db 532 SMKPEIPQOLNGSDCGMFTCKYADYISRDKITFTQHOMPLFRKKMWWEILHQOL 587

Search completed: December 16, 2002, 20:11:15
Job time : 773 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 19:26:20 ; Search time 167 Seconds
(without alignments)
33.125 Million cell updates/sec

Title: US-09-848-852A-3

Perfect score: 1772

Sequence: 1 MYSQRFWTGIVARRGAHLA.....DMPKLRRIQIKELCHCKLTIV 332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105981 seqs, 16662342 residues

Total number of hits satisfying chosen parameters: 105981

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1772	100.0	332	10	US-09-848-852A-3
2	222.5	12.6	621	10	US-09-856-247A-2
3	122.5	6.9	212	12	US-10-008-461-2
4	90.5	5.1	427	10	US-09-815-242-13640
5	89.5	5.1	427	10	US-09-815-242-13443
6	86	4.9	824	10	US-09-866-582-34
7	84	4.7	304	10	US-09-925-302-614
8	82.5	4.7	674	10	US-09-765-272-200
9	82	4.6	803	10	US-09-801-368-394
10	80	4.5	439	9	US-09-890-813-16
11	80	4.5	865	10	US-09-815-242-11364
12	79	4.5	638	10	US-09-815-242-5222
13	79	4.5	642	10	US-09-815-242-12143
14	78.5	4.4	782	9	US-10-041-007-16
15	78.5	4.4	782	10	US-09-887-586A-46
16	78.5	4.4	782	10	US-09-903-012-46
17	78.5	4.4	817	9	US-10-041-007-15
18	78.5	4.4	820	10	US-09-866-582-35
19	78.5	4.4	1503	9	US-10-007-706-1

20	78	4.4	453	10	US-09-905-173-27	Sequence 27, Appl
21	77.5	4.4	381	10	US-09-815-242-4912	Sequence 4912, Ap
22	77.5	4.4	406	10	US-09-815-242-10686	Sequence 10686, A
23	77	4.3	4563	9	US-09-870-759-128	Sequence 128, App
24	76.5	4.3	524	10	US-09-815-242-11557	Sequence 11557, A
25	76.5	4.3	591	10	US-09-887-586A-24	Sequence 24, Appl
26	76.5	4.3	591	10	US-09-903-012-24	Sequence 24, Appl
27	76	4.3	1620	10	US-09-815-242-5136	Sequence 5126, Ap
28	75.5	4.3	1167	10	US-09-815-242-11522	Sequence 11522, A
29	75	4.2	756	10	US-09-745-763-142	Sequence 142, App
30	74.5	4.2	1073	10	US-09-819-249-2	Sequence 2, Appl
31	74.5	4.2	1150	10	US-09-946-239-9	Sequence 9, Appl
32	74	4.2	510	9	US-09-991-496-50	Sequence 50, Appl
33	74	4.2	510	10	US-09-874-923-50	Sequence 7, Appl
34	74	4.2	543	10	US-09-938-956-7	Sequence 7, Appl
35	74	4.2	548	10	US-09-887-586A-52	Sequence 52, Appl
36	74	4.2	548	10	US-09-903-012-52	Sequence 52, Appl
37	74	4.2	599	10	US-09-887-586A-22	Sequence 22, Appl
38	74	4.2	599	10	US-09-903-012-22	Sequence 22, Appl
39	74	4.2	1415	10	US-09-815-242-11036	Sequence 11036, A
40	73.5	4.1	288	10	US-09-925-297-617	Sequence 617, App
41	73.5	4.1	782	10	US-09-815-242-12984	Sequence 12984, A
42	73.5	4.1	983	12	US-10-047-757-2	Sequence 2, Appl
43	73	4.1	386	10	US-09-881-752A-184	Sequence 184, App
44	73	4.1	448	10	US-09-815-242-10596	Sequence 10596, A
45	73	4.1	865	10	US-09-815-242-11536	Sequence 11536, A

ALIGNMENTS

RESULT 1
US-09-848-852A-3
; Sequence 3, Application US/09848852A
; Patent No. US20020106373A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; Tang, Y. Tom
; Corley, Neil C.
; Guegler, Karl J.
; Yue, Henry
; Patterson, Chandra
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH EGF SIGNALING
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/848,852A
; APPLICATION NUMBER: 07-May-2001
; FILING DATE: 07-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/069,725
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0515 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BEPINOT01
CLONE: 2056178
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-848-852A-3

Query Match 100.0%; Score 1772; DB 10; Length 332;
Best Local Similarity 100.0%; Pred. No. 3.6e-167;
Matches 332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSQAQFWCTIWARGAHLAPPDASILISNVCSIGDHVAQELFQSGDLGMAEAEAPGPK 60
DB 1 MYSQAQFWCTIWARGAHLAPPDASILISNVCSIGDHVAQELFQSGDLGMAEAEAPGPK 60

QY 61 AQHSPLEHVTVCQSILDEFLQYGSILPLSTDEVEVEKLEDFIQOEFSTPSRGLVLQ 120
DB 61 AQHSPLEHVTVCQSILDEFLQYGSILPLSTDEVEVEKLEDFIQOEFSTPSRGLVLQ 120

QY 121 LIQSYORMPGNAWGRFVAYKRHLVTMDLGLTLYGQNLNDQVMNMGDLVMDTVPEKV 180
DB 121 LIQSYORMPGNAWGRFVAYKRHLVTMDLGLTLYGQNLNDQVMNMGDLVMDTVPEKV 180

QY 181 HFNSFFYDKLRTKGYDGVRKWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFD 240
DB 181 HFNSFFYDKLRTKGYDGVRKWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFD 240

QY 241 SQTLLNRCPKHIAYLQAEVKKORLDPHQWGKYFKMVARQNNDSCGAFVLYQCKH 300
DB 241 SQTLLNRCPKHIAYLQAEVKKORLDPHQWGKYFKMVARQNNDSCGAFVLYQCKH 300

QY 301 LALSOPFSFTQDMPKLRQIYKELCHCKLT 332
DB 301 LALSOPFSFTQDMPKLRQIYKELCHCKLT 332

RESULT 2
US-09-856-247A-2
; Sequence 2, Application US/09856247A
; Patent No. US20020151028A1
; GENERAL INFORMATION:
; APPLICANT: Lima, Christopher
; APPLICANT: Mossesova, Elena
; TITLE OF INVENTION: Structure-Based Drug Design for Ulpl Protease Substrates
; FILE REFERENCE: 2650/1G681-US1
; CURRENT APPLICATION NUMBER: US/09/856, 247A
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 60/205, 336
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-856-247A-2

Query Match 12.6%; Score 222.5; DB 10; Length 621;
Best Local Similarity 25.2%; Pred. No. 6.5e-14;
Matches 77; Conservative 56; Mismatches 120; Indels 53; Gaps 12;

QY 36 DHVAQEL-FQSGDLGMAEAEAPGKAGQHSPLREHVTVCQSILDEFLQYGSILPLST 94
DB 355 DYLNOKLAFDRSILEFKDFKRYNEILLERKKIQEDLKKKQELAKK-----KLVP--- 405

QY 95 DEVEKLEDFIQOEFSTPSRGLVLQLQSYORMPGNAWGRFVAYKRHLVTMDLGLT 154
DB 406 -ELNEKDDQVQKALA--SRENTQMLNRDNIE-----ITVRDEKTL 443

QY 155 YQCNWLNDQVMNMGDLVMDTVPEKVHFNFSFFYDKLRTKGYDGVRKWTK-----NVDLFP 210
DB 444 APRRLWLDNTIIEFFMKYTEKSTPNTV-AFNSFFYTNLSERGQYGVRRWVRWKRKKTKQIDKLD 502

QY 211 KELLIPILH-EVHWSLISVDVRRRTITYFDS-----QRTLLNRCPKHIAYLQAEAVKDD 265
DB 503 K--IETPINLQSHWALGIIDLKKTIGYVDSLSNGPNAMSFALITDLQKYVMEESKHTI 560

QY 266 RLDFHQWGKYFKMVARQNNDSCGAFVLYQCKHLSQPSFTQDMPKLRQIYKEL 325
DB 561 GDFD-----DLIHLDCPOQPGYDCGIYVCMNTLYGSADAPLDYDKAIRMRRTI---- 611

QY 326 CHCKLT 331
DB 612 AHLILT 617

RESULT 3
US-10-008-461-2
; Sequence 2, Application US/10008461
; Patent No. US20020127692A1
; GENERAL INFORMATION:
; APPLICANT: Barbara Ink
; APPLICANT: Alan Lewis
; TITLE OF INVENTION: New Protein
; FILE REFERENCE: QGI034(p80209)
; CURRENT APPLICATION NUMBER: US/10/008,461
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 0027905.9
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-008-461-2

Query Match 6.9%; Score 122.5; DB 12; Length 212;
Best Local Similarity 23.9%; Pred. No. 0.0001;
Matches 47; Conservative 32; Mismatches 75; Indels 43; Gaps 7;

QY 139 VAYKRHLVTMDLGLTLYGQNLNDQVMNMGDLVMDTVPEKVHFNFSFFYDKLR----- 192
DB 6 LSYMDSLLRQSDVSLDDPPSWLNDHIIGFAFE-----YFANSQFHDCSDHVSFIS 55

QY 193 -----TKGYDGVKRWTKNVDIFNKELLIPILHLE-----VHWSLISVDVRRRTI 236
DB 56 PEVTOFIKTSNPALIAMFLEPLDLPNKRYYFLAINDSNQAGGTHWSLLVYLQDKNSF 115

QY 237 TYFDSQRTLLNRCPKHIAYLQAEVKK-DRLDHFHQWGKYFKMVARQNNDSCGAFVL 295
DB 116 PHYDSHSRSNSVHAKQVAEKLEAFGLGRGDKLAF-----VEEKAPAQNSYDCGMVVI 168

QY 296 QYCKHLSQPSFTQQ 312
DB 169 --CNTEALCQNF-FRQQ 182

RESULT 4
US-09-815-242-13640
; Sequence 13640, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242

```

RESULT 5
US-09-815-242-13443
; Sequence 13443, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

```

```

RESULT 6
US-09-866-582-34
; Sequence 34, Application US/09866582
; Patent No. US20020127620A1
; GENERAL INFORMATION:
; APPLICANT: Witman, George B.
; APPLICANT: Pazour, Gregory J.
; APPLICANT: Rosenbaum, Joel L.
; APPLICANT: Cole, Douglas G.
; TITLE OF INVENTION: INTRAFACELLULAR TRANSPORT
; FILE REFERENCE: 07917-145001
; CURRENT APPLICATION NUMBER: US/09/866,582
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/206,923
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 824
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-866-582-34

```

	Query Match	4.9%;	Score 86;	DB 10;	Length 824;	
	Best Local Similarity	17.6%;	Pred. No. 2.9;			
	Matches	47;	Conservative	48;	Mismatches	98; Indels 74; Gaps 10;
Qy	67	LREEHVTCVQSILDEEFTQTGSLPLSTDEWVEKLEIDFQQEFSPSRKGVLVLQLIOSYQ	126			
	: : : :	: : : : : : :	: : : : : : :			
Db	526	LTYEKLNRDLBALDCFLLKHLAIL--RNSAEVLYQIANYIEL-MENPSQ--AIEMDMQVVS	580			
Qy	127	RMPGNAMVRGFRAVKRRHVLTMDDGLTLYGONWLNDQVMNMYGD--LVMDTPTEPKVHFHN	184			
	: : : :	: : : : : : :	: : : : : : :			

Db 581 VIPTDPQV-----LSKLGELYDREGDKSQAFYYYSYKYPNVIEWLWG 627
QY 185 SFYDKLRTKGYGVKRWTKNDIENKELLPIHLEHVHWSLISVDVRRRTITY---FDS 241
Db 628 AYIID-----TQFWKAIQYFPRASLIQP--TQVKWQLMVASCFFRRSGNYOKALDT 676
QY 242 ORTLNRRCPKHAQYLAQAEAVKADRDLDFHQGWGYFKMNVARQNNDSDCGAFVLOQYCKHL 301
Db 677 YKDTHRKFFENV-----ECLRLFLVRLCTDL 701
QY 302 ALSQPSFTQ-----ODMPKLRRIQYK 323
Db 702 GLKDAOEYARKLRLKLEKMEIREQRIK 728

RESULT 7

US-09-925-302-614
; Sequence 614, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 614
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (208)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (229)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-614

Query Match 4.7%; Score 84; DB 10; Length 304;

Best Local Similarity 18.7%; Pred. No. 1.1;
Matches 61; Conservative 52; Mismatches 99; Indels 114; Gaps 16;

QY 65 SPLREHVTVCQVSIIDFELQTYGSLIPSTDEWVEKLEDFQOEFSTPSRKGLVLQLIQS 124
Db 3 NPWEKTQET-VORIL---LEPKYLLQLPGKQVTKLSQAFNHWLKVPEDK---LQIIIE 55
QY 125 YORMPGNAMV-----RGFRVAYKRHLVTMDLGTLYGQNLNDQVMNMGDILV 172
Db 56 VTEMLHNASLLDIEDNSKLRRGPPVAH-----SIYG----- 88
QY 173 MDTVPKVKHFFNSFY---DKLRTKGY-DGVRWTKNVDIPNKKELLPIH-----LEVHW 224
Db 89 ---IPSVINSANYVFLGLEKLVTLTDHPDAVKLFTRO-----LLELHOGQGLDIYW 136
QY 225 S-----LISVDVRRRTITYEDSQRTL-----NRRCP 250
Db 137 RNYTCPTPEEYKAMVLQKTGGLFGLAVGLMQLFSDYKEDLPLNLTGLFFQIIRDDVAN 196
QY 251 KHIQYLAQAYKKDRLD-----FHQGW---KGYFKMNVARQNNDSDCGAFVLOQYCKH 300
Db 197 LHSKEYSENKSCEDLTGEGKSFPTIHAIWSRXSTQVQNILQRTEH---IDIKKYCVH 253
QY 301 -LALSQPSFTQODMPKLRRIQYKEL 325
Db 254 YLEDVGSFSEYTRNTLKELEKAYKQI 279

RESULT 8

US-09-765-272-200
; Sequence 200, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-09-765-272-200

Query Match 4.7%; Score 82.5; DB 10; Length 674;

Best Local Similarity 22.1%; Pred. No. 4.8;
Matches 56; Conservative 35; Mismatches 79; Indels 83; Gaps 15;

QY 53 EAERPERAGOHSPURE---EHVTCVQS---ILDEFLOT-YGSLIPLSTDEVV----- 98
Db 44 EOENAGLSAARTGLNNMSGNVITFVDSDDWIEQDYVETLYKKIYEQADIAGVNYYSFN 103
QY 99 -----EKLED---IFQOEFSTPSRKGLVLQLIQSYQRMPGNAMVGRFRV 139
Db 104 ESEGMFYPHILGDSYKRYVDNVNVSIFENLYET-----QEMKSFALISAWCK 149
QY 140 AKYRHLVLM---DDLTLGQNLNDQVMNMGDILVDPVKVHFHNSFFYDKLRTKGYD 197
Db 150 LYLKARLEQLRFDIGKLGEDGYLNQKV-----LLSEKVIYLNKSLYAIRKGG-- 198
QY 198 GVKR-WTKN-----VDIENKELLLI-----PI--HLEVHWSLISVDVRR-----RTIT 237
Db 199 SUSRVWTEKWHMALVDAMSERITLLANMGYPLEKHLAVIRQMLVSLANGQASGLSDTAT 258
QY 238 YFD---SQTLMNR 247
Db 259 YKEFEMKQRLNLQ 271

RESULT 9

US-09-801-368-394
; Sequence 394, Application US/09801368
; Patent No. US20020128250A1

```

; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 394
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-394

Query Match          4.5%; Score 82; DB 10; Length 803;
Best Local Similarity 22.2%; Pred. No. 6.8;
Matches 56; Conservative 41; Mismatches 95; Indels 60; Gaps 12;

Qy  50 MAEAEERGERAGQHSPLREH-----HVTVCQSILDEFLOTQYGSILPLSTDEVVE 99
Db  528 MVSRGDYDQGRNGAKKRSQLLKNPPTSLINDVQNLNLSISKIDYEN---ETVOYNE 583
Qy  100 KLEDF-----QOEFSTPSRGL--VLQIOSYQRMPCGNM-----VRGERVAYKR 143
Db  584 KLEKLHKLNFQRELANSRQLANVKQKDEYSILMQQLNLNKGIEEBEESFREESKK 643
Qy  144 HVLTMDDIGTYGQNLNDQVMNMYGDLVMDT--VPEKVHFFNFFYDKLRTKGYDGVKRW 202
Db  644 LGIIADESS---GIDWDSSE-----YDADEPKVEFLSDFLEDKIQ--KNYEG--DI 688
Qy  203 TKNVDIFNKELLPIHLEVHWSLI-----SDVRRRTITYFDSORTLNRCPCPKHIAKY 256
Db  689 SKLLEAESKEQIMEQIRNQLPAEKIQSMLPPTVLLKARINAY-----KRNDKHLTNV 740
Qy  257 LQAEAVKKDRLD 268
Db  741 LDTISTKQSELE 752

RESULT 10
US-09-890-813-16
; Sequence 16, Application US/09890813
; Publication No. US20020183486A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; TITLE OF INVENTION: Aspartate Kinase
; FILE REFERENCE: BB1430 PCT
; CURRENT APPLICATION NUMBER: US/09/890,813
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/172944
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Triticum aestivum

Query Match          4.5%; Score 80; DB 10; Length 865;
Best Local Similarity 23.7%; Pred. No. 12;
Matches 36; Conservative 27; Mismatches 51; Indels 38; Gaps 6;

Qy  60 KAGQHSPLREEHVTCVQSILDEFLOTQYGSILPLS-----TDEVVKEKIDFQOEFSTPSR 114
    ||| ||| ||| : : : ||| | : ||| : : :

```

```

US-09-890-813-16

Query Match          4.5%; Score 80; DB 9; Length 439;
Best Local Similarity 21.0%; Pred. No. 4.6;
Matches 60; Conservative 52; Mismatches 86; Indels 88; Gaps 18;

Qy  45 GSDLGMAEEARPGEKAGQHSPLREHV-----TCVQSILDEFLOTQYGSILP---LS 93
    ||| ||| | : ||| | ||| : ||| : ||| :
Db  157 GSDL---TATTICKALG---LREIQVWKDVGDLTCDPNI-----YANAVVPVYLT 201
Qy  94 TDEVVEKL---EDIFQOEFSTPSRK-GLVLQLIQSYQR-MEGNAM--VRGFR-----VA 140
    ||| | : : ||| : : ||| : ||| : ||| :
Db  202 FDEAELAYFGAQLHPQSMRPARREGGIPVRKNSYHRAHPTVITKTRDMRKSLTISIV 261
Qy  141 YKRHVLTMDLIGT-LYGONLNDQVMNMYGDLVMDTVPEKVHFFNFFYDKLRTK----- 194
    ||| : ||| : ||| : ||| : ||| : ||| :
Db  262 LKSNITMLDIDVSTRMLGQYGFIAKVFISFDLGTISV-----DSVATSEVSIS 308
Qy  195 -GYDGVKRWTKNVDIFNKEL-----LLPIHLEVHWSLISV--DVRRRTITYFDSQR 243
    ||| : ||| : ||| : ||| : ||| : ||| :
Db  309 LTLDPCKLWSR--ELIQOELDHVVVELEKIAVWHLLQHRSLISLIGNVQRSSLL--EK 363
Qy  244 TLNRRCPKHIAKYLOAEAVKKDRLDHFQGWKGYPKMNVARQNDS 289
    | : : : : : : : : : : : : : : :
Db  364 AFN-----VLRNGVNVQVISQGASKVNISLVVNDSE 395

RESULT 11
US-09-815-242-11364
; Sequence 11364, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11364
; LENGTH: 865
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11364

Query Match          4.5%; Score 80; DB 10; Length 865;
Best Local Similarity 23.7%; Pred. No. 12;
Matches 36; Conservative 27; Mismatches 51; Indels 38; Gaps 6;

Qy  60 KAGQHSPLREEHVTCVQSILDEFLOTQYGSILPLS-----TDEVVKEKIDFQOEFSTPSR 114
    ||| ||| ||| : : : ||| | : ||| : : :

```

Db 653 KAFDHNLSEELLGUKNLKEDFNHVSLEDLKKASPIENFVAEKLKSDYENK----- 706

QY 115 KGLVLQIQSYQRMPCGNMVRGRVAY-----KRHLVTMDLGT-----LYGQWLNQD 163

Db 707 -----MKVLQSEQRS-----RIRIYVLQILDNAWEHLTYTMDLKTGTLNRCYNQKDP 756

QY 164 V-----MNVGLVMDVTPEKVFHFNFFYD 189

Db 757 VEYKESYNLFEFIEDIKTEAITSKIQFE 788

RESULT 12

US-09-815-242-5222

; Sequence 5222, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5222

; LENGTH: 638

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-5222

Query Match 4.5%; Score 79; DB 10; Length 638;

Best Local Similarity 18.7%; Pred. No. 9.8;

Matches 56; Conservative 49; Mismatches 84; Indels 110; Gaps 13;

QY 30 NVCSIGD-----HVAQLFGQSDLGMAEEAERPGEKAGQ-HSPLEP----- 70

Db 59 NVSKIKDLRLGYLTQMTFNSNATVFEEMSKPFEHKRMESLIKEETDWSKHANDYDSD 118

QY 71 ----HVTVCQSILDEFQYGLSLPLSTDEVVEKLEDFQOEFSTP-----SRKGL 117

Db 119 TVKTHMSRYESLSNQPELEGQYQESKIKTVLHGL-NFSEEDFNKPINDFSGGQKTRL 177

QY 118 VLQLIQSYQRMPCGNMVRGRVAYKRHLVTMD-----DLGTYLQWLNQWMMNYGDL 171

Db 178 AQMLN-----EPDLLLDEPTNHLDET-----TKWLEDYLRVFKGAI 216

QY 172 VMDTPEKVHFFNSFFYDKLRTKGYD-----GVKRWTKNVDIFNKELLIPIHLEHWSLI 227

Db 217 VIIS-----HDRVFLDKIVTQYDVALGDVKRYGVNYEFP----- 251

QY 228 SVDVRRRTITY-----FDSQ-----RTLNRCPKHIAKYLOAEAVKKDRLD 268

Db 252 ----IQORDLYYOKRMOEYESQQAETKRLTFFVEKNITRASTSGMAKSRKILEKMERID 307

Db 252 ----IQORDLYYOKRMOEYESQQAETKRLTFFVEKNITRASTSGMAKSRKILEKMERID 307

RESULT 13

US-09-815-242-12143

; Sequence 12143, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12143

; LENGTH: 642

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-12143

Query Match 4.5%; Score 79; DB 10; Length 642;

Best Local Similarity 18.7%; Pred. No. 9.9;

Matches 56; Conservative 49; Mismatches 84; Indels 110; Gaps 13;

QY 30 NVCSIGD-----HVAQLFGQSDLGMAEEAERPGEKAGQ-HSPLEP----- 70

Db 59 NVSKIKDLRLGYLTQMTFNSNATVFEEMSKPFEHKRMESLIKEETDWSKHANDYDSD 118

QY 71 ----HVTVCQSILDEFQYGLSLPLSTDEVVEKLEDFQOEFSTP-----SRKGL 117

Db 119 TVKTHMSRYESLSNQPELEGQYQESKIKTVLHGL-NFSEEDFNKPINDFSGGQKTRL 177

QY 118 VLQLIQSYQRMPCGNMVRGRVAYKRHLVTMD-----DLGTYLQWLNQWMMNYGDL 171

Db 178 AQMLN-----EPDLLLDEPTNHLDET-----TKWLEDYLRVFKGAI 216

QY 172 VMDTPEKVHFFNSFFYDKLRTKGYD-----GVKRWTKNVDIFNKELLIPIHLEHWSLI 227

Db 217 VIIS-----HDRVFLDKIVTQYDVALGDVKRYGVNYEFP----- 251

QY 228 SVDVRRRTITY-----FDSQ-----RTLNRCPKHIAKYLOAEAVKKDRLD 268

Db 252 ----IQORDLYYOKRMOEYESQQAETKRLTFFVEKNITRASTSGMAKSRKILEKMERID 307

RESULT 14

US-10-041-007-16

; Sequence 16, Application US/10041007

; Patent No. US20020164736A1

; GENERAL INFORMATION:

; APPLICANT: Matsuda, Seichi P.T.
; APPLICANT: Schepmann, Hala G
; TITLE OF INVENTION: Ginkgo Biloba Levopimaradiene Synthase
; FILE REFERENCE: P02081US1
; CURRENT APPLICATION NUMBER: US/10/041,007
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259,881
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 782
; TYPE: PRT
; ORGANISM: Abies grandis
US-10-041-007-16

Query Match 4.4%; Score 78.5; DB 9; Length 782;
Best Local Similarity 23.4%; Pred. No. 15;
Matches 40; Conservative 32; Mismatches 54; Indels 45; Gaps 9;
QY 43 FQSGDLGMAEEA-ERPGE-----KAGQHSPLREE-----HVTCVQSILDEFLOTYGSIL- 89
DB 482 FRDSGLPLFTFARERPLEFYLVAAGTYEPQYAKCRFLFTKVACIQTVLDDMDYDTYGTLD 541
QY 90 -IPLSTDEV-----VEKLEDFQOEFSTPSRKGVLQILQISYQRMPCGNAMVRGFRVAY 141
DB 542 ELKLFTEAVRRWDLSTENLPDYMKLCYQI--YYDIVHEVAWEAEKEQGRELVSPFRKGW 599
QY 142 KRVL-----TWDDLGLTYGQNW-----LNDQVMNMYGDLVMD 174
DB 600 EDYLLGYVEAEWLAAYVPTLDE----YIKNGITSIGORILLSGVLMD 646

RESULT 15
US-09-887-586A-46
; Sequence 46, Application US/09887586A
; Patent No. US20020094556A1
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. US20020094556A11, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/887,586A
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 782
; TYPE: PRT
; ORGANISM: Abies grandis
US-09-887-586A-46

Query Match 4.4%; Score 78.5; DB 10; Length 782;
Best Local Similarity 23.4%; Pred. No. 15;
Matches 40; Conservative 32; Mismatches 54; Indels 45; Gaps 9;
QY 43 FQSGDLGMAEEA-ERPGE-----KAGQHSPLREE-----HVTCVQSILDEFLOTYGSIL- 89
DB 482 FRDSGLPLFTFARERPLEFYLVAAGTYEPQYAKCRFLFTKVACIQTVLDDMDYDTYGTLD 541
QY 90 -IPLSTDEV-----VEKLEDFQOEFSTPSRKGVLQILQISYQRMPCGNAMVRGFRVAY 141
DB 542 ELKLFTEAVRRWDLSTENLPDYMKLCYQI--YYDIVHEVAWEAEKEQGRELVSPFRKGW 599
QY 142 KRVL-----TWDDLGLTYGQNW-----LNDQVMNMYGDLVMD 174

Db 600 EDYLLGYVEAEWLAAYVPTLDE----YIKNGITSIGORILLSGVLMD 646
Search completed: December 16, 2002, 19:39:09
Job time : 168 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 17:51:04 ; Search time 37 Seconds
(without alignments)
264.011 Million cell updates/sec

Title: US-09-848-852A-3
Perfect score: 1772
Sequence: 1 MYSQRFWGTIARRGAHLA.....DMPKLRRIYKELCHCKLTV 332

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91.5	5.2	654	US-09-090-793-2	Sequence 2, Appl
2	91.5	5.2	661	US-08-375-709-3	Sequence 3, Appl
3	91.5	5.2	661	US-08-752-929-3	Sequence 3, Appl
4	86	4.9	727	US-09-134-001C-4067	Sequence 4067, Ap
5	83	4.7	1151	US-09-134-001C-3242	Sequence 3242, Ap
6	82.5	4.7	343	US-08-180-209B-56	Sequence 56, Appl
7	82.5	4.7	343	US-08-474-853-56	Sequence 56, Appl
8	82.5	4.7	343	US-09-166-205B-56	Sequence 56, Appl
9	82.5	4.7	343	PCT-US94-02629-56	Sequence 56, Appl
10	82.5	4.7	451	US-09-357-251-35	Sequence 35, Appl
11	82.5	4.7	674	US-08-961-083-200	Sequence 200, App
12	82	4.6	353	US-08-034-650-11	Sequence 11, Appl
13	82	4.6	353	US-08-449-015-11	Sequence 11, Appl
14	81.5	4.6	500	US-08-117-083-68	Sequence 68, Appl
15	81.5	4.6	633	US-09-041-991A-10	Sequence 10, Appl
16	81	4.6	407	US-08-989-370-6	Sequence 6, Appl
17	81	4.6	668	US-09-134-001C-4816	Sequence 4816, Ap
18	81	4.6	2987	US-08-970-269A-29	Sequence 29, Appl
19	81	4.6	2987	US-09-407-562-29	Sequence 29, Appl
20	81	4.6	3959	US-08-970-269A-30	Sequence 30, Appl
21	81	4.6	3959	US-09-407-562-30	Sequence 30, Appl
22	80.5	4.5	449	US-09-134-001C-5673	Sequence 5673, Ap
23	79.5	4.5	335	US-09-057-762-7	Sequence 7, Appl
24	79.5	4.5	335	US-08-326-119A-7	Sequence 7, Appl
25	79.5	4.5	885	US-09-342-648-9	Sequence 9, Appl
26	79.5	4.5	1010	US-09-134-001C-5178	Sequence 5178, Ap
27	78.5	4.4	436	5405943-4	Patent No. 5405943

ALIGNMENTS

RESULT 1

US-09-090-793-2
; Sequence 2, Application US/09090793
; Patent No. 6140486
; GENERAL INFORMATION:
; APPLICANT: Calgene, LLC
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
; FILE OF INVENTION: of polyketide-like synthesis genes in plants
; FILE REFERENCE: CGNE.131.01US
; CURRENT APPLICATION NUMBER: US/09/090,793
; CURRENT FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-09-090-793-2

Query Match	5.2%	Score 91.5;	DB 4;	Length 654;	
Best Local Similarity	21.2%	Pred. No. 0.5;			
Matches	78;	Conservative	43;	Mismatches 118; Indels 129; Gaps 21;	
QY	1	MYS---	AORFWGTIARRGAHLAPPDASILISNVCSIGDHVAQELFGQSDLG----	MAEE 53	
Db	171	IYSHSHADHFGGA----	RGVQEMFPDVVKVYGS-----	DNITKEIVDENVLGNAMSRA 220	
QY	54	AERPEKAGQH-----	SPLREHVTCVQSILDEFLOTGSLIPLSTDEVVEKLED 103		
Db	221	AYVGATGLKHHGHDVAALGKLSGEITYVAP--	DYTLNSEGKWETLT----	IDGLEM 274	
QY	104	IFQOEEST-----	PSRKGVLVLQIQSYQRMFG-----	NAMV 134	
Db	275	VFMDSAGTEASESEMITYIPSKKALWTAEI-	TYOQMHNIYTLRGAKVRDALKWSKDINEMI 333		
QY	135	RGF-----	RVAYKRHV-----	LTMDDLGLTYGQ--	NWLNDQV--MNNYGLDM 173
Db	334	NAGQDVEVLFASHAPVWGQAINDFLRQDNYGLVHNQTLRLANDGVIGQIDGDAIQ 393			
QY	174	DTVPEKVHFFNSFFYDKLRKTYGDKVKNVND-	IPNKELLIIPIHLEHVHWSLISVDVR 232		
Db	394	DTIPESI-----	YKTHWTNGYHGT--	YSHNAKAVYNYL-----	425
QY	233	RRITTFDSQTLNRCP--	KHIKAYLQ-----	AEAVKKDRLDFHQWKGYYF-----	KMN 280
Db	426	----	GYFDMNPANLNPLPTKQESAKFVEYMGGAADAIKRAKDDYAQGEYRFVATALNKVV 481		
QY	281	VARQNDS 288			

Db 482 MAEPENDS 489

RESULT 2

US-08-375-709-3

; Sequence 3, Application US/08375709

; Patent No. 5683898

; GENERAL INFORMATION:

; APPLICANT: YAZAWA, Kazunaga

; APPLICANT: YAMADA, Akiko

; APPLICANT: KATO, Seishi

; APPLICANT: KONDO, Kiyosi

; TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid

; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of

; TITLE OF INVENTION: Eicosapentaenoic Acid

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/375,709

; FILING DATE: 20-JAN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/178,251

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 4-147945

; FILING DATE: 15-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: WEGNER, Harold C.

; REGISTRATION NUMBER: 25,258

; REFERENCE/DOCKET NUMBER: 53466/150/AAOK

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 661 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-375-709-3

Query Match 5.2%; Score 91.5; DB 1; Length 661;

Best Local Similarity 21.2%; Pred. No. 0.51;

Matches 78; Conservative 43; Mismatches 118; Indels 129; Gaps 21;

QY 1 MYS---AQRFWGTIWARGAHLAPPDASILISNVCSIGHVAQELFGSDLG-----MAEE 53

Db 173 IYSHSHADHFGA---RGVQEMFDDVKVYGS-----DNITKEIVDENVLGNAMSRA 222

QY 54 AERPGEKAGQH-----SPLREHVTCVQSILDBFLQTYGSLIPLSTDEVVEKLED 103

Db 223 AYQYGATLGGKHGIVDAALGKLSKEITYVAP--DYTLNSEGKWETLT----IDGLEM 276

QY 104 IFQEEFST-----PSRKGVLQLIQSVQRMGP-----NANV 134

Db 277 VFMDASGTEAESEMIYIPSKALMTAEL-TYOGMHNITYTLGAKVRDALKWSKDINEMI 335

QY 135 RGF-----RVAYKRHY-----LTMDLLGTLYGQ--NWLNDQV--MNMVYGDLYM 173

Db 336 NAFQDVEVLFASHAPVWGNQAINDFLRLQDNYGLVHNQTLRLANDGVGIQDIDAIO 395

QY 174 DTVPEKVHFFNSFFYDKLRTKGYGVKRWTKNVD--IPNKELLIPILHLEVHWSLISVDVR 232

Db 396 DTIPESI-----YKTWHTNGYHGT--YSHNAKAVYNYKL----- 427

QY 233 RRTITYFDSQRTLMRRCP--KHIAKYLO-----APAVKKDRDLFHQGWKGYF-----KMN 280

Db 428 ---GYFDMNPANLPLPTKQESAKFVEYMGADAAIKRAKDDYAQGEYRFVATALNKVV 483

QY 281 VARONNDS 288

Db 484 MAEPENDS 491

RESULT 3

US-08-752-929-3

; Sequence 3, Application US/08752929

; Patent No. 5798259

; GENERAL INFORMATION:

; APPLICANT: YAZAWA, Kazunaga

; APPLICANT: YAMADA, Akiko

; APPLICANT: KATO, Seishi

; APPLICANT: KONDO, Kiyosi

; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing

; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic

; TITLE OF INVENTION: Acid

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/752,929

; FILING DATE: 20-NOV-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/375,709

; FILING DATE: 20-JAN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/178,251

; FILING DATE: 14-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 4-147945

; FILING DATE: 15-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: WEGNER, Harold C.

; REGISTRATION NUMBER: 25,258

; REFERENCE/DOCKET NUMBER: 53466/150/AAOK

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 661 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-752-929-3

Query Match 5.2%; Score 91.5; DB 1; Length 661;

Best Local Similarity 21.2%; Pred. No. 0.51;

Matches 78; Conservative 43; Mismatches 118; Indels 129; Gaps 21;

;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/180,209B
;/ FILING DATE: 11-JAN-1994
;/ CLASSIFICATION: 435
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/031,400
;/ FILING DATE: 11-MAR-1993
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Jackson Esq., David A.
;/ REGISTRATION NUMBER: 26,742
;/ REFERENCE/DOCKET NUMBER: 600-1-074 CIP
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 201 487-5800
;/ TELEFAX: 201 343-1684
;/ TELEX: 133521
;/ INFORMATION FOR SEQ ID NO: 56:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 343 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: peptide
;/ HYPOTHETICAL: NO
;/ ANTI-SENSE: NO
;/ US-08-180-209B-56

Query Match 4.7%; Score 82.5; DB 1; Length 343;
Best Local Similarity 19.5%; Pred. No. 1.7; Mismatches 43; Indels 77; Gaps 13;
Matches 46; Conservative 43; Mismatches 43; Indels 77; Gaps 13;
QY 131 NAMVGRFVAYK-----RHVLTMDDLGTLG--QNWLN-----DOVMNMY-----168
Db 4 NKTVREFNVYNNVPTFMCHKYGRLFEVSEKYGILQNMWDMKFRGEIATLYDPGMFPALL 63
QY 169 ----GDLV-----MDTVPEKVHFNSEFFYDKLRTKGYDGV-----KRWTKNVDIFNK 211
Db 64 KDPNGNVARNGVGPQLGNLTQKHLQVFRDHLINQIPDKSPGPGVGVDFESWR-----115
QY 212 ELLLPITHLEVHWS-----LISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQAEAVKDK 265
Db 116 ----PIFRQ--NWASLOPYKKLSVEVVRHEHFWDDQD--VEQEAKRFRFEKYQ-----161
QY 266 RLDFHOGWKGYFKMVARQNNDSCGAFVLQYCKHLALSQPF-----SFTQQDMPKL 317
Db 162 -LFMEETLKAARMPA-----ANWGYAYPYCYNLTNPQPSAQCEATTMQENDKM 211

RESULT 7
US-08-474-853-56
; Sequence 56, Application US/08474853
; Patent No. 6287559
; GENERAL INFORMATION:
; APPLICANT: King, Te-Piao
; TITLE OF INVENTION: CLONING AND RECOMBINANT PRODUCTION OF
; TITLE OF INVENTION: VESPID VENOM ENZYMES, SUCH AS PHOSPHOLIPASE AND
; TITLE OF INVENTION: HYALURONIDASE, AND IMMUNOLOGICAL THERAPIES BASED THEREON
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,853
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435

;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/180,209
;/ FILING DATE: 11-JAN-1994
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/031,400
;/ FILING DATE: 11-MAR-1993
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Jackson Esq., David A.
;/ REGISTRATION NUMBER: 26,742
;/ REFERENCE/DOCKET NUMBER: 600-1-074 CIPB
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 201 487-5800
;/ TELEFAX: 201 343-1684
;/ TELEX: 133521
;/ INFORMATION FOR SEQ ID NO: 56:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 343 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: peptide
;/ HYPOTHETICAL: NO
;/ ANTI-SENSE: NO
;/ US-08-474-853-56

Query Match 4.7%; Score 82.5; DB 4; Length 343;
Best Local Similarity 19.5%; Pred. No. 1.7; Mismatches 43; Indels 77; Gaps 13;
Matches 46; Conservative 43; Mismatches 43; Indels 77; Gaps 13;
QY 131 NAMVGRFVAYK-----RHVLTMDDLGTLG--QNWLN-----DOVMNMY-----168
Db 4 NKTVREFNVYNNVPTFMCHKYGRLFEVSEKYGILQNMWDMKFRGEIATLYDPGMFPALL 63
QY 169 ----GDLV-----MDTVPEKVHFNSEFFYDKLRTKGYDGV-----KRWTKNVDIFNK 211
Db 64 KDPNGNVARNGVGPQLGNLTQKHLQVFRDHLINQIPDKSPGPGVGVDFESWR-----115
QY 212 ELLLPITHLEVHWS-----LISVDVRRRTITYFDSQRTLNRRCPKHIAKYLQAEAVKDK 265
Db 116 ----PIFRQ--NWASLOPYKKLSVEVVRHEHFWDDQD--VEQEAKRFRFEKYQ-----161
QY 266 RLDFHOGWKGYFKMVARQNNDSCGAFVLQYCKHLALSQPF-----SFTQQDMPKL 317
Db 162 -LFMEETLKAARMPA-----ANWGYAYPYCYNLTNPQPSAQCEATTMQENDKM 211

RESULT 8
US-09-166-205B-56
; Sequence 56, Application US/09166205B
; Patent No. 6372471
; GENERAL INFORMATION:
; APPLICANT: Te Piao KING
; TITLE OF INVENTION: CLONING AND RECOMBINANT PRODUCTION OF VESPID VENOM ENZYMES,
; TITLE OF INVENTION: SUCH AS PHOSPHOLIPASE AND HYALURONIDASE, AND IMMUNOLOGICAL
; TITLE OF INVENTION: THERAPIES BASED THEREON
; FILE OF INVENTION: 2313/0F138US
; CURRENT APPLICATION NUMBER: US/09/166,205B
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Apis mellifera
; US-09-166-205B-56

Query Match 4.7%; Score 82.5; DB 4; Length 343;
Best Local Similarity 19.5%; Pred. No. 1.7; Mismatches 43; Indels 77; Gaps 13;
Matches 46; Conservative 43; Mismatches 43; Indels 77; Gaps 13;
QY 131 NAMVGRFVAYK-----RHVLTMDDLGTLG--QNWLN-----DOVMNMY-----168
Db 4 NKTVREFNVYNNVPTFMCHKYGRLFEVSEKYGILQNMWDMKFRGEIATLYDPGMFPALL 63

[illegible]

```

Db      64 KDPNGNVARNQVPGVQLGNTKTKHLQVFRDHLINQIPDKSFPGVGVIDFESWR----- 115-
|::|| : : : | : : : | : : : |
Qy      212 ELLIPIHLEVHWS-----LISDVRRRTITYFDSQRTLNRRCPKHIARYLQAEEAVKKD 265
| : : | : : | : : | : : | : : | : : |
Db      116 ----PIFRQ-NWASLOPYKKLSVEVVRRHPFDQOR-VEQAKRREKYQG----- 161
| : : | : : | : : | : : | : : | : : |
Qy      266 RLDPHQGWKGFKMVARQNNDSDCAFLVLYCKYCHIALSQPF-----SFTQQDDMPKL 317
| : : | : : | : : | : : | : : | : : |
Db      162 -LFMEETLTAAKRMRPA-----ANWGYYAYPYCYNLTPNQPSAQCEATTMOENDKM 211
| : : | : : | : : | : : | : : | : : |

RESULT 10
US-09-357-251-35
; Sequence 35, Application US/09357251
; Patent NO. 6271441
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Schwaber, James S.
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
; FILE REFERENCE: BB-1193
; CURRENT APPLICATION NUMBER: US/09/357,251
; CURRENT FILING DATE: 1999-07-20
; EARLIER APPLICATION NUMBER: 60/093,530
; EARLIER FILING DATE: July 21, 1998
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-357-251-35

```

COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 200:
SEQUENCE CHARACTERISTICS:
LENGTH: 674 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-200

Query Match 4.7%; Score 82.5; DB 4; Length 674;
Best Local Similarity 22.1%; Pred. No. 4.9;
Matches 56; Conservative 35; Mismatches 79; Indels 83; Gaps 15;
QY 53 EAEPEGEKAGQSPURE---EHVTCVQS---ILDEFLOT-YGSLIPLSTDEVV----- 98
Db 44 EQENAGLSAARTGLNMGSGYVTFVDSDDTIEQDYVETLYKRIYEQADIAVGNYSFN 103
QY 99 -----EKLED---IFQOEFSTPSRKGLVLQLIQSVQRMPGNMYRGERV 139
Db 104 ESEMEFYHILGDSYKERYDNVSIFENDYET-----QEMKSFALISAMGK 149
QY 140 AYKRHRVLTN--DOLGTYLQGNMLNDQVMNMYGDLVMDTVPEKVFHFNSFFYDKLRTKGYD 197
Db 150 LYKARLEQRLFDKLGEDGYLNQVY-----LLSEKVIYLNKSLYAIRKG-- 198
QY 198 GVKR-WTKN-----VDIFNKELLLI-----PI--HLEVHMSLISVDVRR-----RTIT 237
Db 199 SLRSVWTERKMHVLDAMSERITLLANMGYPLEKHLAVYRQMLEVSLANGQASGLSDTAT 258
QY 238 YFD---SORTLNR 247
Db 259 YKEFEMKQRLNQ 271

RESULT 12
US-08-034-650-11
; Sequence 11, Application US/08034650
; Patent No. 5641671
; GENERAL INFORMATION:
; APPLICANT: BOS, Jannetje W.
; APPLICANT: FRENKEN, Leon G.
; APPLICANT: VERRIPS, Cornelis T.
; APPLICANT: VISSER, Christiaan
; TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE
; TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1615 L. Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA

ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/034,650
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/727,235
FILING DATE: 03-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kokulis, Paul N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: PNK/5970/91731
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 353 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-034-650-11
Query Match 4.6%; Score 82; DB 1; Length 353;
Best Local Similarity 20.8%; Pred. No. 2;
Matches 48; Conservative 37; Mismatches 94; Indels 52; Gaps 9;
QY 46 SDLGMAEAEERPEGEKAGQHSPL-----RBEHVTCVQSILDEFLOTYGSILPLSTDEVV 98
Db 69 AEAPMPLPALPGALAGSHAPRLPLAAGRLARTRAVREFFDYCLTAQOGLTPAALDALV 128
QY 99 EKLEIDFQOEFSTPSRKGL-VLQLIQSV-----QRMPGNMYRGERVAYKRHRVLTMDLGT 153
Db 129 RR-ETAAQLDGSQAQAEALGVWRRYRAYFDALAQIPLGDAVLGDKLDPAAQMLAQDRAA 187
QY 154 LYQGNLNDQVMNMYGDLVMDTVPEKVFHFNSFFYDKLRTKGYDGVKRWTKNVDIFNKEL 213
Db 188 -----LADRTLGEAE-----PFFGDEQRRQRHD-----LERIRLAND-- 220
QY 214 LLPIHLEVHMSLISVDVRRRTITYFDSQRLNRRCPKHIKYLQAEAVKK 264
Db 221 -----TTLSPEQKAARLAALDAQLTDPDERA-QOAAALHAQODAVTK 259

RESULT 13
US-08-449-015-11
; Sequence 11, Application US/08449015
; Patent No. 5804409
; GENERAL INFORMATION:
; APPLICANT: BOS, Jannetje W.
; APPLICANT: FRENKEN, Leon G.
; APPLICANT: VERRIPS, Cornelis T.
; APPLICANT: VISSER, Christiaan
; TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE
; TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1615 L. Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/449,015
;; FILING DATE: 24-MAY-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/727,235
;; FILING DATE: 03-JUL-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kokulis, Paul N.
;; REGISTRATION NUMBER: 16,773
;; REFERENCE/DOCKET NUMBER: PNK/5970/91731
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 861-3000
;; TELEFAX: (202) 822-0944
;; TELEX: 6714627 CUSH
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 353 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-449-015-11

Query Match 4.6%; Score 82; DB 1; Length 353;
Best Local Similarity 20.8%; Pred. No. 2;
Matches 48; Conservative 37; Mismatches 94; Indels 52; Gaps 9;
QY 46 SLDGMAEAEERCEKAGQHSPL-----REEHVTCSQSIDLEFLQYGLSLPLSTDEVV 98
DB 69 AEAAMPPLPAALPGALAGSHAPRLPLAAGGLARTRAVREFDYCLTAQELTPAALDALV 128
QY 99 EKLEDFQEEFTSPSKGI-VLQIQSY----QRMGNAVGRFVAYKRHVLTWDLLGT 153
DB 129 RR-ETAAQLDGSPAQAELGVWRRYRAYFDALQAQPGDVLGDKLDPAAQLADQRAA 187
QY 154 LYQGNLNDQVNMNYGDLVMDTVPEKVHFFNSFFYDKLRTKGYDGVKRWTKNVDIFNKEL 213
DB 188 -----LADRTLGEAAE-----PFFGDEQRQRHD-----LERIRIAND-- 220
QY 214 LLIPHTLHVHNSLISVDVRRRTITFYDSORTLNRRCPKHIAKYLOAEAVKK 264
DB 221 -----TTLSPEQKAARLAALDAQLTPDERA-QQAALHAQQDAVTK 259

RESULT 14
US-08-117-083-68
; Sequence 68, Application US/08117083
; Patent No. 5719054
; GENERAL INFORMATION:
; APPLICANT: Boursnell, Michael E.
; APPLICANT: Inglis, Stephen C.
; APPLICANT: Munro, Alan J.
; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human
; TITLE OF INVENTION: Papilloma Virus Proteins
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Walter H. Dreger
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,083
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

;; NAME: Dreger, Walter H.
;; REGISTRATION NUMBER: 24,190
;; REFERENCE/DOCKET NUMBER: A-58783
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-781-1989
;; TELEFAX: 415-398-3249
;; TELEX: 910 277299
;; INFORMATION FOR SEQ ID NO: 68:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 500 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FEATURE:
;; NAME/KEY: Protein
;; LOCATION: 1..500
;; OTHER INFORMATION: /note= "Xaa refers to stop codon in
;; OTHER INFORMATION: the open reading frame."
US-08-117-083-68

Query Match 4.6%; Score 81.5; DB 1; Length 500;
Best Local Similarity 19.8%; Pred. No. 4;
Matches 51; Conservative 37; Mismatches 86; Indels 83; Gaps 9;
QY 82 FLQTYGSLPLSTDEVVEKLEDFQ-----QEFSTP-----SRKGLVLQLIQS 124
DB 64 YNKTGSCIQDMSBALEYLSLKSGGWSRPPEQFEYDPGVEDTESIERLVEEFFNR 123
QY 125 YQMPGNAVGRFVAYKRHVLTMDLGLTYGQNLNDQVNMNYGDLVMDTVPEKVHFFN 184
DB 124 SELQAGESVKFGNSINVKHTSVSAKQLRTIRIQ-----LPLYSHLLP----- 166
QY 185 SFYDKLRTKGYDGVKRWTKNVDIFNKELLIPQ-HLEVHNSL----- 226
DB 167 -----TQRYDICSLELIITHTXLDLSRWYMTTSDWXRMIKXSISKDF 208
QY 227 ---ISVDVRRRTITFYDSORTLN--RRCPKH---IAKYLOAEAVKKDRLDHOGW----- 273
DB 209 VLCISVRRKRTSSRVSDSSRYINLGMRLPRHTCALSKWKDAVVLCIQGNHGIWLAVS 268
QY 274 KGYFKMNAVARQNNSDC 290
DB 269 RSLQMGNGCHIXNFNC 285

RESULT 15
US-09-041-991A-10
; Sequence 10, Application US/09041991A
; Patent No. 6107278
; GENERAL INFORMATION:
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Muller-Cohn, Judy
; TITLE OF INVENTION: Toxins Active Against Pests
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,991A
; FILING DATE: 13-MAR-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:

NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA-709
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 633 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-041-991A-10

Query Match 4.6%; Score 81.5; DB 3; Length 633;
Best Local Similarity 23.2%; Pred. No. 5.7;
Matches 52; Conservative 33; Mismatches 96; Indels 43; Gaps 11;
QY 24 ASILISNVCS-IGDHVAQEL---FGSDLGMAEEAEPGEKAGOHSPLEEHVTCVQSI 78
DB 58 ASFLKKGSLGKRLSELRLNLIFFSGSTNLMEDILRETEKF-LNOKLNTDLSRVNAE 116
QY 79 L-----DEFLQYGSLLPLSTDEVVEKLEDFQOEFSTPSRKGLVLQLIQS 124
DB 117 LTGLQANVEEFNRQVDNFLNPNRNAVPLSTSSVTMQQLFLNRLSQFQMGVQLLL--- 173
QY 125 YQRMPCGNMVRGFRVAYKRHV-LTMDDLG-----TLYG-QNWLNDQVMNMGDLVMDTVE 178
DB 174 ---LPLFAQAANTHLSYIRDVILNAEEWGISAATLRTYQNHLENYTRD-YSNYCIDTYQT 229
QY 179 KVHFENSEFYDKLRTKGY-----DGVKRWTKNVDIFNKELLLI 216
DB 230 AFRGLNRIHDMLEFRYMFNLNVFEYVSIWS-----LFKYQSLLV 269

Search completed; December 16, 2002, 19:28:42
Job time : 39 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 19:28:00 ; Search time 899 Seconds
(without alignments)
49.209 Million cell updates/sec

Title: US-09-848-852A-3

Perfect score: 1772

Sequence: 1 MYSQRFWGTWIRRGHAHLA.....DMPKLRQYKELCHCKLV 332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1667	94.1	371	21	AAV50911 Human fetal brain
2	1667	94.1	371	22	AAV50911 Human fetal brain
3	1667	94.1	438	22	AAW41055 Human polypeptide
4	1667	94.1	438	22	AAW41056 Human polypeptide
5	1667	94.1	568	22	AAW41056 Human polypeptide
6	1558.5	88.0	354	22	AAW39270 Amino acid sequenc
7	1470	83.0	309	23	ABP41139 Human polypeptide
8	781.5	44.1	270	22	ABW25617 Human ovarian anti
9	614.5	34.7	119	22	AAO04440 Human protein sequ
10	490	27.7	191	22	AAW92493 Human polypeptide

11	456	25.7	643	22	AAW31973 Amino acid sequenc
12	393	22.2	509	22	AAW31976 Amino acid sequenc
13	393	22.2	589	22	AAW4011 Human stomach canc
14	393	22.2	589	22	AAW95254 Human protein sequ
15	311	17.6	468	22	ABW59432 Drosophila melanog
16	307	17.3	1513	22	ABW63592 Human secreted pro
17	280	15.8	51	21	AAW34355 Arabidopsis thalia
18	260.5	14.7	446	21	AAW45771 Arabidopsis thalia
19	260.5	14.7	489	21	AAW45772 Arabidopsis thalia
20	258.5	14.6	305	21	AAW45772 Arabidopsis thalia
21	257.5	14.5	446	21	AAW25055 Arabidopsis thalia
22	257.5	14.5	489	21	AAW25054 Arabidopsis thalia
23	255.5	14.4	305	21	AAW25056 Arabidopsis thalia
24	249.5	14.1	198	20	AAW86186 Human EST product
25	249.5	14.1	489	20	AAW86184 Arabidopsis ESD4 (
26	205	11.6	209	20	AAW86185 Rice EST product s
27	201.5	11.4	566	23	ABW90943 Herbicidally activ
28	190	10.7	548	21	AAW77935 A. thaliana enviro
29	169.5	9.6	1148	22	ABW64875 Drosophila melanog
30	158.5	8.9	98	22	AAW01755 Human gene 11 enco
31	154	8.7	560	22	ABW71015 Drosophila melanog
32	143	8.1	362	21	AAW91846 Human secreted pro
33	143	8.1	1112	22	AAW78297 Human SUMO-1 SEQ 1
34	143	8.1	1112	22	AAW40358 Human polypeptide
35	143	8.1	1115	22	ABG14039 Novel human diagno
36	125	7.1	368	22	ABW94888 Human protein sequ
37	122.5	6.9	238	23	ABW97801 Human secretory po
38	119.5	6.7	223	23	ABW04392 Human protein phos
39	116	6.5	162	22	ABG07019 Novel human diagno
40	112	6.3	51	21	AAW34354 Gene 4 human secre
41	110	6.2	215	22	ABW63890 Drosophila melanog
42	107.5	6.1	272	23	ABW53442 Lactococcus lactis
43	106	6.0	2464	22	AAW78919 Human protein SEQ
44	104.5	5.9	272	23	ABW54137 Lactococcus lactis
45	104.5	5.9	272	23	ABW55429 Lactococcus lactis

ALIGNMENTS

RESULT 1

AAV50911

ID AAV50911 standard; Protein; 371 AA.

XX AAV50911;

AC AAV50911;

XX 10-MAR-2000 (first entry)

XX Human fetal brain cDNA clone vb7_1 derived protein #1.

XX Human; secreted protein; treatment: nutritional activity; cytokine;
XX cell proliferation; cell differentiation; hematopoiesis regulation;
XX tissue growth; activin; inhibitor; chemotactic; chemokinetic; hemostatic;
XX thrombolytic; anti-inflammatory; invasion suppressor; tumor inhibition;
XX gene therapy.

OS Homo sapiens.

XX WO9955721-A1.

PN 04-NOV-1999.

XX 23-APR-1999; 99WO-US08504.

XX 24-APR-1998; 98US-0082904.

PR 11-JUN-1998; 98US-0088994.

PR 12-JUN-1998; 98US-0089278.

PR 02-JUL-1998; 98US-0091647.

PR 24-AUG-1998; 98US-0097639.

PR 22-APR-1999; 99US-0097639.

PA (ALPH-) ALPHAGENE INC.

XX

PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
XX WPI: 2000-052801/04.
DR N-PSDB; AA243782.
XX
XX New polynucleotides encoding secreted human proteins, derived from
PT human fetal brain, adult skin, adult brain, adult heart, adult thymus
PT and adult aorta cDNA libraries.
XX
XX Claim 21a; Page 223-225; 282pp; English.
XX
XX This invention describes novel human secreted proteins which are encoded
CC by polynucleotides obtained from fetal brain, adult skin, adult brain,
CC adult heart, adult thymus and adult aorta cDNA libraries. The
CC polynucleotides and proteins are predicted to have biological activities
CC which would make them suitable for treating, preventing or ameliorating
CC medical conditions in humans and animals, although no supporting data
CC is given. Suggested activities include nutritional activity, cytokine
CC and cell proliferation/differentiation activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, hematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, hemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cacharin/tumor
CC invasion suppressor activity, and tumor inhibition activity. The
CC polynucleotides are also stated to be useful for gene therapy.
CC AA50905-Y50947 represent the secreted proteins described in the method
CC of the invention which are encoded by the polynucleotides represented in
CC AA243777-243808.
XX
XX Sequence 371 AA;
SQ
Query Match 94.1%; Score 1667; DB 21; Length 371;
Best Local Similarity 100.0%; Pred. No. 7e-157;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGQHSPLREHVTVCVOSI 78
DB 58 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGQHSPLREHVTVCVOSI 117
QY 79 LDEFLOTYGSLLPLSTDVEVEKLEDFQOEFSTPSRKGVLVLQILOSORMPGNAMVRGFR 138
DB 118 LDEFLOTYGSLLPLSTDVEVEKLEDFQOEFSTPSRKGVLVLQILOSORMPGNAMVRGFR 177
QY 139 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 198
DB 178 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 237
QY 199 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 258
DB 238 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 297
QY 259 AEAVKKDLRDLFHQGWKGYFKMVARQNNDSDCGAFVLOYCKHLALSOPFSFTQDMPKLR 318
DB 298 AEAVKKDLRDLFHQGWKGYFKMVARQNNDSDCGAFVLOYCKHLALSOPFSFTQDMPKLR 357
QY 319 RQIYKELCHCKLTV 332
DB 358 RQIYKELCHCKLTV 371
RESULT 2
ID AAM39269
XX AAM39269 standard; Protein; 371 AA.
XX
XX AAM39269;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2414.
XX
KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW anyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX Homo sapiens.
XX
XX WO200153312-A1.
PN
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000WO-US34263.
PF
XX
XX 21-JAN-2000; 2000US-0488725.
PR
XX 25-APR-2000; 2000US-0552317.
PR
XX 09-JUL-2000; 2000US-0598042.
PR
XX 19-JUL-2000; 2000US-0620312.
PR
XX 03-AUG-2000; 2000US-0653450.
PR
XX 14-SEP-2000; 2000US-0662191.
PR
XX 19-OCT-2000; 2000US-0693036.
PR
XX 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AA158425.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 2414; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 371 AA;
SQ
Query Match 94.1%; Score 1667; DB 22; Length 371;
Best Local Similarity 100.0%; Pred. No. 7e-157;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGQHSPLREHVTVCVOSI 78
DB 58 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGQHSPLREHVTVCVOSI 117
QY 79 LDEFLOTYGSLLPLSTDVEVEKLEDFQOEFSTPSRKGVLVLQILOSORMPGNAMVRGFR 138
DB 118 LDEFLOTYGSLLPLSTDVEVEKLEDFQOEFSTPSRKGVLVLQILOSORMPGNAMVRGFR 177
QY 139 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 198
DB 178 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 237
QY 199 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 258
DB 238 VAYKRHLVTMDLGLTYGQNLNDQVMNMGDLVMDTPVEKVFHFNFSFYDKLRTKGYDG 297

QY	259	AEAVKKDR	LDFHQGWKGYFKMVARQNNDSCGAFVLOYCKHLALSQPFSTQODMPKLR	318
Db	298	AEAVKKDR	LDFHQGWKGYFKMVARQNNDSCGAFVLOYCKHLALSQPFSTQODMPKLR	357
QY	319	ROIYKELC	HKCKLTV	332
Db	358	ROIYKELC	HKCKLTV	371
RESULT	3			
AAM41055				
ID	AA041055	standard; Protein; 438 AA.		
XX	AC	AA041055;		
XX	DT	22-OCT-2001 (first entry)		
XX	DE	Human polypeptide SEQ ID NO 5986.		
XX	XX			
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;			
KW	peripheral nervous system; neuropathy; central nervous system; CNS;			
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;			
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;			
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;			
KW	leukaemia.			
XX	XX			
OS	Homo sapiens.			
XX	XX			
PN	WO200153312-A1.			
XX	XX			
PD	26-JUL-2001.			
XX	XX			
PF	26-DEC-2000; 2000WO-US34263.			
XX	XX			
PR	21-JAN-2000; 2000US-0488725.			
PR	25-APR-2000; 2000US-0552317.			
PR	09-JUL-2000; 2000US-0598042.			
PR	19-JUL-2000; 2000US-0620312.			
PR	03-AUG-2000; 2000US-0653450.			
PR	14-SEP-2000; 2000US-0662191.			
PR	19-OCT-2000; 2000US-0693036.			
PR	29-NOV-2000; 2000US-0727344.			
XX	XX			
PA	(HYSE-) HYSEQ INC.			
XX	XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;			
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;			
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;			
XX	XX			
DR	WPI; 2001-442253/47.			
DR	N-PSDB; AA160211.			
XX	XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders			
PT	such as central nervous system injuries -			
XX	XX			
PS	Example 2; SEQ ID NO 5986; 10078pp; English.			
XX	XX			
CC	The invention relates to human nucleic acids (AA157798-AA161369) and			
CC	the encoded polypeptides (AA038642-AA042213) with nootropic,			
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful			
CC	in gene therapy. A composition containing a polypeptide or polynucleotide			
CC	of the invention may be used to treat diseases of the peripheral nervous			
CC	system, such as peripheral nervous injuries, peripheral neuropathy and			
CC	localised neuropathies and central nervous system diseases, such as			
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic			
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the			
CC	utilisation of the activities such as: Immune system suppression,			
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic			
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,			
CC	assays for receptor activity, arthritis and inflammation, leukaemia and			
CC	C.N.S disorders.			
CC	CC			
CC	Note: The sequence data for this patent did not form part of the printed			

DR WPI; 2001-442253/47.
DR N-PSDB; AAI60212.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 5987; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 438 AA;

Query Match 94.1%; Score 1667; DB 22; Length 438;
Best Local Similarity 100.0%; Pred. No. 8.9e-157;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGHSPLEEHVTCVQSI 78
DB 125 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGHSPLEEHVTCVQSI 184
QY 79 LDEFLOTYGSILPLSTDEVVEKLEDFQOEFSTPSRKGLVLQIQSYQRMPCGNAMVRGFR 138
DB 185 LDEFLOTYGSILPLSTDEVVEKLEDFQOEFSTPSRKGLVLQIQSYQRMPCGNAMVRGFR 244
QY 139 VAYKRHLVTMDLGLTYGONLNDQVMNMGDLVMDTVPEKVFHFNFSFYDKLRTKGYDG 198
DB 245 VAYKRHLVTMDLGLTYGONLNDQVMNMGDLVMDTVPEKVFHFNFSFYDKLRTKGYDG 304
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYEDSQRTLNRRCPKHIAYLQ 258
DB 305 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYEDSQRTLNRRCPKHIAYLQ 364
QY 259 AEAVKKDLRDFHQGWKGYFKMNVARQNNDSCGAFVLOYCKHLALSQPFSTQODMPKLR 318
DB 365 AEAVKKDLRDFHQGWKGYFKMNVARQNNDSCGAFVLOYCKHLALSQPFSTQODMPKLR 424
QY 319 ROIYKELCHCKLTV 332
DB 425 ROIYKELCHCKLTV 438

RESULT 5
AAB31977
ID AAB31977 standard; Protein; 568 AA.
XX
AC AAB31977;
XX
DT 15-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human SENP3 polypeptide.
XX
KW Human; de-sentrinase; SENP; SENP2; SENP3; chromosome 12q13.1;
KW sentrin-specific protease; sentrin; anti-proliferative agent;
KW anti-viral agent; herpes simplex-1; cytomegalovirus; sentrinisation;
KW PML; tumour suppressor; acute promyelocytic leukaemia.
OS Homo sapiens.
XX

PN WO200109292-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-US20884.
XX
PR 31-JUL-1999; 99US-0146774.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Gong L, Yeh ETH;
XX
DR WPI; 2001-182947/18.
DR N-PSDB; AAF54831.
XX
PT New polynucleotide encoding de-sentrinase polypeptides, useful as
XX antiproliferative or antiviral agents -
XX
XX Example 6; Page 1121-122; 122pp; English.
XX
CC The present sequence represents human de-sentrinase (SENP) 3 polypeptide.
CC The specification describes SENP1 and SENP2 polypeptides. The SENP1
CC gene is found in chromosome 12q13.1. SENP1 is found in the nucleus,
CC SENP3 is found in the nucleolus, and SENP3 is found in both locations.
CC SENP polypeptides are sentrin-specific proteases that remove sentrin
CC from some sentrinised peptides, but do not affect proteins modified by
CC ubiquitin or NEDD8. SENP polypeptides are used to identify specific
CC modulators of SENP. These modulators are potential anti-proliferative
CC and anti-viral agents such as herpes simplex-1 or cytomegalovirus.
CC SENP polypeptides are used for studying the role of sentrinisation in
CC the biological function of PML, a tumour suppressor implicated in
CC development of acute promyelocytic leukaemia. Fragments of SENP
CC polynucleotides are used as hybridisation probes and amplification
CC primers for detecting gene expression or preparing mutated sequences,
CC also as antisense sequences for inhibiting SENP expression.
XX
SQ Sequence 568 AA;

Query Match 94.1%; Score 1667; DB 22; Length 568;
Best Local Similarity 100.0%; Pred. No. 1.3e-156;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGHSPLEEHVTCVQSI 78
DB 255 LAPPDASILISNVCSIGDHVAQELFGQSDLGMAEAEAPGKAGHSPLEEHVTCVQSI 314
QY 79 LDEFLOTYGSILPLSTDEVVEKLEDFQOEFSTPSRKGLVLQIQSYQRMPCGNAMVRGFR 138
DB 315 LDEFLOTYGSILPLSTDEVVEKLEDFQOEFSTPSRKGLVLQIQSYQRMPCGNAMVRGFR 374
QY 139 VAYKRHLVTMDLGLTYGONLNDQVMNMGDLVMDTVPEKVFHFNFSFYDKLRTKGYDG 198
DB 375 VAYKRHLVTMDLGLTYGONLNDQVMNMGDLVMDTVPEKVFHFNFSFYDKLRTKGYDG 434
QY 199 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYEDSQRTLNRRCPKHIAYLQ 258
DB 435 VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYEDSQRTLNRRCPKHIAYLQ 494
QY 259 AEAVKKDLRDFHQGWKGYFKMNVARQNNDSCGAFVLOYCKHLALSQPFSTQODMPKLR 318
DB 495 AEAVKKDLRDFHQGWKGYFKMNVARQNNDSCGAFVLOYCKHLALSQPFSTQODMPKLR 554
QY 319 ROIYKELCHCKLTV 332
DB 555 ROIYKELCHCKLTV 568

RESULT 6
AAM39270
ID AAM39270 standard; Protein; 354 AA.
XX
AC AAM39270;
XX

DT 22-OCT-2001 (first entry)
 XX Human polypeptide SEQ ID NO 2415.
 DE
 XX
 XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 XX Homo sapiens...
 OS
 XX WO200153312-A1.
 XX
 XX 26-JUL-2001.
 XX
 XX 26-DEC-2000; 2000WO-US34263.
 XX
 XX 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 XX 09-JUL-2000; 2000US-0598042.
 XX 19-JUL-2000; 2000US-0620312.
 XX 03-AUG-2000; 2000US-0653450.
 XX 14-SEP-2000; 2000US-0662191.
 XX 19-OCT-2000; 2000US-0693036.
 XX 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI58426.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 XX Example 4; SEQ ID NO 2415; 10078pp; English.
 PS
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 XX Sequence 354 AA;
 SQ
 Query Match 88.0%; Score 1558.5; DB 22; Length 354;
 Best Local Similarity 94.6%; Pred. No. 3.9e-146;
 Matches 29; Conservative 0; Mismatches 0; Indels 17; Gaps 1;
 19 LAPPDASILINVCISGDHVAQELFQSGDLGMAEAEAPGKAGQHSPLREHVTVCQSI 78
 DB LAPPDASILINVCISGDHVAQELFQSGDLGMAEAEAPGKAGQHSPLREHVTVCQSI 117
 79 LDEFLOTQYGLIPLSTDRVWEKLEDFIQQEFSTPSRKLVLQLITQSYORMPGNAWGR 138
 DB LDEFLOTQYGLIPLSTDRVWEKLEDFIQQEFSTPSRKLVLQLITQSYORMPGNAWGR 177

QY 139 VAYKRHLVTMDLGLTXQGNWLNDOVMNMYGDLVMDTVPEKVFHFNSEFFYDKLRTKYD 198
 DB 178 VAYKRHLVTMDLGLTXQGNWLNDOVMNMYGDLVMDTVPEKVFHFNSEFFYDKLRTKYD 237
 QY 199 VKRWTKNVDIPNKKELLIPHLEVHWSLISVDVRRRTITYPDSQRTLNRRCPKHIAYLQ 258
 DB 238 VKRWTKNVDIPNKKELLIPHLEVHWSLISVDVRRRTITYPDSQRTLNRRCPKHIAYLQ 297
 QY 259 AEAVKDRDLDPHQCKGKGYEKNVARNQNDSCGAFVLQYCKHLSOPFSFTQODMPKLR 318
 DB 298 AEAVKDRDLDPHQCKGKGYEKNVARNQNDSCGAFVLQYCKHLSOPFSFTQODMPKLR 340
 QY 319 ROIYKELCHCKLTV 332
 DB 341 ROIYKELCHCKLTV 354
 RESULT 7
 ABP41139
 ID ABP41139 standard; Protein; 309 AA.
 AC ABP41139;
 XX
 XX 22-AUG-2002 (first entry)
 XX Human ovarian antigen HOF027, SEQ ID NO:2271.
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 XX Homo sapiens.
 OS
 XX WO200200677-A1.
 XX
 XX 03-JAN-2002.
 XX
 XX 07-JUN-2001; 2001WO-US18569.
 XX
 XX 07-JUN-2000; 2000US-209467P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Birse CE, Rosen CA;
 XX WPI; 2002-147878/19.
 DR N-PSDB; ABQ54216.
 XX
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 XX Claim 11; SEQ ID NO 2271; 2922pp; English.
 PS
 XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,

CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 309 AA;

Query Match 83.0%; Score 1470; DB 23; Length 309;
Best Local Similarity 97.2%; Pred. No. 2e-137;
Matches 276; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 49 GMAEAEAPRGKAGHSPLEHEHVTVCQSILDEFLOTYSGLPLSTDEVVEKLEDFQOE 108
DB 26 GHRRGRAGEKAGHSPLEHEHVTVCQSILDEFLOTYSGLPLSTDEVVEKLEDFQOE 85
QY 109 FSTPSKGLVLQIQSYQRMPCNMGVFRVAYKRHLVTMDLGLTLYGQNLNDQVMNMY 168
DB 86 FSTPSKGLVLQIQSYQRMPCNMGVFRVAYKRHLVTMDLGLTLYGQNLNDQVMNMY 145
QY 169 GDLYVMDTPEKVFHFNFSFYDKLRTPGVDGVRKWTKNVDIFNKELLPIHLEVHWSLIS 228
DB 146 GDLYVMDTPEKVFHFNFSFYDKLRTPGVDGVRKWTKNVDIFNKELLPIHLEVHWSLIS 205
QY 229 VDVRRTITYFDSQTLNRCPKHIAKYLAQAVKRDLDHFQGWKGYFKMNVARQNDS 288
DB 206 VDVRRTITYFDSQTLNRCPKHIAKYLAQAVKRDLDHFQGWKGYFKMNVARQNDS 265
QY 289 DCGAFVLYQCKHALSQPSFSTQDMPKLRROIYKELCHCKLTV 332
DB 266 DCGAFVLYQCKHALSQPSFSTQDMPKLRROIYKELCHCKLTV 309

RESULT 8
AAM25617
ID AAM25617 standard; Protein: 270 AA.
XX
AC AAM25617;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:1132.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antitumor; cardiovascular; antianaemic; anaemia;
KW antiagregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; diabetes; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.
XX
OS Homo sapiens.

XX WO200153455-A2.
PN 26-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-US35017.
XX
PR 23-DEC-1999; 99US-0471275.
XX
PR 21-JAN-2000; 2000US-0488725.
XX
PR 25-APR-2000; 2000US-0552317.
XX
PA (HYSE-) HYSEQ INC.
XX
Tang YT, Liu C, Drmanac RT;
XX
WPI; 2001-457603/49.
XX
DR N-PSDB; AAH999558.
XX
Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX Claim 20; Page 235; 1217pp; English.
XX
XX AAM259166 to AAM99904 encode the human proteins given in AAM25225 to
XX AAM25963. The proteins can have activities based on the tissues and
XX cells they are expressed in, such as: antinflammatory; antirheumatic;
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
XX central nervous system; virucide; anti-HIV; fungicide; antitumor;
XX cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;
XX antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX encoding them can be used in gene therapy, antisense therapy and vaccine
XX production. The proteins and polynucleotides are useful for screening for
XX agonists or antagonists of a protein and for the treatment and diagnosis
XX of disorders associated with the activity of a protein e.g. inflammation,
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX osteoporosis, severe combined immunodeficiency, eczema, allergic
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX neurological disorders.
XX
SQ Sequence 270 AA;

Query Match 44.1%; Score 781.5; DB 22; Length 270;
Best Local Similarity 54.0%; Pred. No. 4.2e-69;
Matches 148; Conservative 54; Mismatches 65; Indels 7; Gaps 4;
QY 58 GEKAGHSPLEHEHVT-CVQSILDEFLOTYSGLPLSTDEVVEKLEDFQOEFTPSRKG 116
DB 1 GKNQKASPDVDEQLSVCLSGFLDEVMKKYGLVPLSEKEVLGLRKLQVFNEDFS--NRKP 58
QY 117 LVLLQIQSYQRMPCNMGVFRVAYKRHLVTMDLGLTLYGQNLNDQVMNMYGDLVMDTV 176
DB 59 FINEITNWRARHOKC---NFRIFNKHMLDMDLTLGQNLNDQVINMYGELIMDAV 115
QY 177 PEKVHFNFSFYDKLRTPGVDGVRKWTKNVDIFNKELLPIHLEVHWSLISVDVRRRTI 236
DB 116 PDKVHFNFSFYDKLRTPGVDGVRKWTKNVDIFNKELLPIHLEVHWSLITVLSNRII 175
QY 237 TYFDSQTLNRCPKHIAKYLAQAVKRDLDHFQGWKGYFKMNVARQNDSGCAFVLQ 296
DB 176 SFYDSQGHFFFCVENIRKYLLEAREKNLNL-QGMOTAVTKCIPOQKNDSDCGVFVLQ 234
QY 297 YCKHALSQPSFSTQDMPKLRROIYKELCHCKL 330
DB 235 YCKLALQKPFQSDMPKLRROIYKELCHCKL 268
RESULT 9

Db 156 LDFLQYGLSLPLSTDEYVEKLEDFQOEFSTPSR 191
|||||

RESULT 11

AAB31973
ID AAB31973 standard; Protein; 643 AA.

AC AAB31973;

DT 15-MAY-2001 (first entry)

DE Amino acid sequence of a human SENP1 polypeptide.

Human; de-sentrinase; SENP; SENP1; SENP2; SENP3; chromosome 12q13.1;
sentrin-specific protease; sentrin; anti-proliferative agent;
anti-viral agent; herpes simplex-1; cytomegalovirus; sentrinisation;
PML; tumour suppressor; acute promyelocytic leukaemia.

OS Homo sapiens.

PN WO200109292-A2.

XX 08-FEB-2001.

XX 31-JUL-2000; 2000WO-US20884.

XX 31-JUL-1999; 99US-0146774.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Gong L, Yeh ETH;

DR WPI: 2001-182947/18.

DR N-PSDB; AAF54827.

PT New polynucleotide encoding de-sentrinase polypeptides, useful as
antiproliferative or antiviral agents -

PS Example 1; Fig 2; 122pp; English.

CC The present sequence represents human de-sentrinase (SENP) 1 polypeptide.
CC The specification describes SENP2 and SENP3 polypeptides. The SENP1
CC gene is found in chromosome 12q13.1. SENP1 is found in the nucleus,
CC SENP3 is found in the nucleolus, and SENP3 is found in both locations.
CC SENP polypeptides are sentrin-specific proteases that remove sentrin
CC from some sentrinised peptides, but do not affect proteins modified by
CC ubiquitin or NEDD8. SENP polypeptides are used to identify specific
CC modulators of SENP. These modulators are potential anti-proliferative
CC and anti-viral agents such as herpes simplex-1 or cytomegalovirus.
CC SENP polypeptides are used for studying the role of sentrinisation in
CC the biological function of PML, a tumour suppressor implicated in
CC development of acute promyelocytic leukaemia. Fragments of SENP
CC polynucleotides are used as hybridisation probes and amplification
CC primers for detecting gene expression or preparing mutated sequences,
CC also as antisense sequences for inhibiting SENP expression.

XX Sequence 643 AA;

Query Match 25.7%; Score 456; DB 22; Length 643;

Best Local Similarity 34.0%; Pred. No. 3.2e-36;

Matches 112; Conservative 60; Mismatches 111; Indels 46; Gaps 11;

OY 24 ASILSNVCSIGDHVAQE---LFQSGDLGMAEEAERPECKAGQHS-----PLR 68

DB 338 AEWLKELTSYDVSRRARERIEQKALALQLOQRQER--EHSVHDSVELHLRVPLE 395

OY 69 EE-HVTCVQSILDEFLQYGLSLPLSTDEYVEKLEDFQOEFSTPSRKGLVLQLIQSYQR 127

DB 396 KEIPVTVVQE-----TQKGKHLTSEDEFFPEITEM-EKEIKNVFRNG----- 438

OY 128 MFGNAWGRGFRVAYKRHHVLTDDGLTLYGQNWLNDOVNMVYGDLYMDTVPEK-----VHFF 183

Db 439 NODEVLSEAFRL-----TITRKDIQTLNHLNWLNDIINFYNNMLMERSKEKGLPSVHAF 493
OY 184 NSFFFYDKLRTKGYDGVRKWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSOR 243
DB 494 NTFEFTKLKTAGYQAVKRWTKKVDVFSVDILLVPHLGVHWC LAVVDVFRKKNTITYDSMG 553
OY 244 TLNRRCPKHIKYLQAEAVKKDRDLDFH-QGWMGYFKMN-VARQNNDSDCGAFVLOYCKHL 301
DB 554 GINNEACRILLQYLKQESIDKKRKEFDTNQWOLFSSKQIPQOMNGSDCGMPACKYADCI 613
OY 302 ALSQPFSTQDMPKLRRIQYKELCHCKL 330
DB 614 TKDRPINFTHQHPYFRKRMVWEILHRKL 642

RESULT 12

AAB31976
ID AAB31976 standard; Protein; 509 AA.

AC AAB31976;

DT 15-MAY-2001 (first entry)

DE Amino acid sequence of a human SENP2 polypeptide.

Human; de-sentrinase; SENP; SENP1; SENP2; SENP3; chromosome 12q13.1;
sentrin-specific protease; sentrin; anti-proliferative agent;
anti-viral agent; herpes simplex-1; cytomegalovirus; sentrinisation;
PML; tumour suppressor; acute promyelocytic leukaemia.

OS Homo sapiens.

PN WO200109292-A2.

XX 08-FEB-2001.

XX 31-JUL-2000; 2000WO-US20884.

XX 31-JUL-1999; 99US-0146774.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Gong L, Yeh ETH;

DR WPI: 2001-182947/18.

DR N-PSDB; AAF54830.

PT New polynucleotide encoding de-sentrinase polypeptides, useful as
antiproliferative or antiviral agents -

PS Example 6; Page 118-120; 122pp; English.

CC The present sequence represents human de-sentrinase (SENP) 2 polypeptide.
CC The specification describes SENP1 and SENP3 polypeptides. The SENP1
CC gene is found in chromosome 12q13.1. SENP1 is found in the nucleus,
CC SENP3 is found in the nucleolus, and SENP3 is found in both locations.
CC SENP polypeptides are sentrin-specific proteases that remove sentrin
CC from some sentrinised peptides, but do not affect proteins modified by
CC ubiquitin or NEDD8. SENP polypeptides are used to identify specific
CC modulators of SENP. These modulators are potential anti-proliferative
CC and anti-viral agents such as herpes simplex-1 or cytomegalovirus.
CC SENP polypeptides are used for studying the role of sentrinisation in
CC the biological function of PML, a tumour suppressor implicated in
CC development of acute promyelocytic leukaemia. Fragments of SENP
CC polynucleotides are used as hybridisation probes and amplification
CC primers for detecting gene expression or preparing mutated sequences,
CC also as antisense sequences for inhibiting SENP expression.

XX Sequence 509 AA;

Query Match 22.2%; Score 393; DB 22; Length 509;

Best Local Similarity 29.6%; Pred. No. 4.2e-30;

Matches 97; Conservative 64; Mismatches 107; Indels 60; Gaps 9;

PS Claim 8; SEQ ID 17423; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

SQ Sequence 589 AA;

Query Match 22.2%; Score 393; DB 22; Length 589;

Best Local Similarity 29.6%; Pred. No. 5.2e-30;

Matches 97; Conservative 64; Mismatches 107; Indels 60; Gaps 9;

QY 13 ARGAHLAPPDASILISVCSIGDVAQELFGQSLGMAEAEERGERAGOHSPLEEHV 72

DB 311 SRGQLEP-----DLSEVSARLRGS-----GNGLLRRR-- 342

QY 73 TCVCQSILDELTQYGLSLPLSTDEYVEKLEDFIQOEFSTPSRKGVLQIQSYQRMPCNA 132

DB 343 --VSIETKEKNCSEKDRRTDDELLETDN-EKEISNALGHGQDEILSS----- 391

QY 133 MYRGFRVAYKRVHLTMDLGLTYGQNLWDQVMNMYGDLVMDTVPEK-----VHFNSEFFY 188

DB 392 -----AFKLRI-TRGDHITLKNYHNLNDEVINFYMSLLTERSEKRSGLVLPATYAINTFVPRLLQAGHAGIKR 443

QY 189 DKLRKGVGKRVTKNVDIENKELLIPILHEVHWSLISVDVRRRTITYFDSQRTLNRR 248

DB 444 PKLSSGGQAVKRVTKGVNLEFQEIILPIHRKVHWSLVVIDLRKKLKYLDMSGCKGHR 503

QY 249 CPKHIAKYLAQAVKDKRLDF-----HGWKGYFKMVARQNNDSDCGAFVLYQCKHLA 302

DB 504 ICEILLVQDESKTKRNSDLNLEWTHHSMKPH---EIPQQLNGSDCGMFTCKYADVIS 560

QY 303 LSQPSFTQDDMPKLRRIYKELCHCKL 330

DB 561 RDKPITFTQHQMPLEFRKKMNVWEILHQOL 588

RESULT 15

ABB59432

ID ABB59432 standard; Protein: 468 AA.

XX

AC ABB59432;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 5088.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

XX

OS Drosophila melanogaster.

XX

FN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL03535.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Disclosure; SEQ ID NO 5088; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 468 AA;

Query Match 17.6%; Score 311; DB 22; Length 468;

Best Local Similarity 28.5%; Pred. No. 5.2e-22;

Matches 70; Conservative 55; Mismatches 85; Indels 36; Gaps 6;

QY 89 LIPLSTDEVVEKLEDFIQOEFSTPSRKGVLQIQSYQRMPCGNAMVRCFRVAYKRHV-LT 147

DB 245 LIPL-TEEHDRFRNEITQDDKST-----EIIKFNLHIT 277

QY 148 MDDLGLTYGQNLWDQVMNMYGDLVMDTVPEK-----VHFNSEFFYDKLRKGVGKRV 201

DB 278 TEDICTFNGKWLNDVINFYMSLLTERSEKRSGLVLPATYAINTFVPRLLQAGHAGIKR 337

QY 202 WTKNVDIPNKKELLIPILHE-VHWSLISVDVRRRTITYFDSQRTLNRRCPKHIAKYLAQAE 260

DB 338 WTRKVDLFSKDIIPVPVHCNGVHWCMAIILRNKTIIRYDSKGPNRVPLDALEKYLREE 397

QY 261 AVKDKRLDFH-QGWKGYFKMVARQNNDSDCGAFVLYQCKHLALSQPSFTQDDMPKLR 319

DB 398 SIFPKKQFDTSDFVIESVQNIPIQLDGDGCGIFSCMFAEYITCDVPIITFTQSEMFLFRK 457

QY 320 QIYKEL 325

DB 458 KMALEI 463

Search completed: December 16, 2002, 19:54:29

Job time : 900 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 16, 2002, 20:11:20 ; Search time 98 Seconds
(without alignments)
1339.725 Million cell updates/sec

Title: US-09-848-852A-3
Perfect score: 1772
Sequence: 1 MYSQRFWTGIWARRGAHLA.....DMPKLRRIYKELCHCKLTV 332

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 355320 seqs, 197730502 residues

Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09848852/runat_16122002_132039_19075/app_query.fasta_1.519
-DB=PublishedApplications_NA -OFMT=fastap -SUFFIX=trnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09848852_@CGN_1.1.49_@runat_16122002_132039_19075
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	1772	100.0	1991	10	US-09-848-852A-4	Sequence 4, Appli
c 2	704	39.7	432	10	US-09-983-965-835	Sequence 835, App
c 3	661.5	37.3	503	10	US-09-777-564-1710	Sequence 1710, Ap
4	222.5	12.6	1866	10	US-09-856-247A-1	Sequence 1, Appli

5	162	9.1	435	10	US-09-867-701-4735	Sequence 4735, Ap
6	144	8.1	5540	10	US-09-960-253-74	Sequence 74, Appl
7	130.5	7.4	264	10	US-09-878-574-7902	Sequence 7902, Ap
8	122.5	6.9	639	12	US-10-008-461-1	Sequence 1, Appli
9	115.5	6.5	100	10	US-09-728-445-223	Sequence 223, App
10	110	6.2	263	10	US-09-294-093B-3098	Sequence 3098, Ap
11	108	6.1	4835	10	US-09-956-004-58	Sequence 58, Appl
c 12	103	5.8	547	10	US-09-924-401-19	Sequence 19, Appl
13	96	5.4	472	10	US-09-783-590-7988	Sequence 7988, Ap
14	93.5	5.3	2529	10	US-09-925-302-171	Sequence 171, App
15	90.5	5.1	1284	10	US-09-815-242-9542	Sequence 139, App
16	90	5.1	2395	9	US-10-063-547-139	Sequence 44, Appl
17	90	5.1	2395	12	US-10-036-342-44	Sequence 139, App
18	90	5.1	2395	12	US-10-006-867-139	Sequence 9345, Ap
19	89.5	5.1	1284	10	US-09-815-242-9345	Sequence 9345, Ap
c 20	89.5	5.1	29793	10	US-09-973-451-38	Sequence 38, Appli
21	87	4.9	5342	10	US-09-852-922-1	Sequence 1, Appli
22	86.5	4.9	1575	10	US-09-815-242-7460	Sequence 7460, Ap
c 23	86	4.9	890	10	US-09-070-927A-606	Sequence 606, App
24	86	4.9	2457	9	US-09-954-531-177	Sequence 177, App
25	86	4.9	2457	9	US-09-954-531-593	Sequence 593, App
26	86	4.9	2457	10	US-09-919-497-42	Sequence 42, Appl
27	84	4.7	9783	10	US-09-728-952-47	Sequence 47, Appl
28	83	4.7	720	10	US-09-910-943-733	Sequence 733, App
29	82.5	4.7	2023	10	US-09-765-272-199	Sequence 199, App
30	82.5	4.7	368004	10	US-09-949-654-3	Sequence 3, Appli
31	82	4.6	2412	10	US-09-801-368-393	Sequence 393, App
32	80	4.5	1658	9	US-09-890-813-15	Sequence 15, Appli
33	80	4.5	2598	10	US-09-815-242-7267	Sequence 7267, Ap
34	80	4.5	3320	12	US-10-047-757-1	Sequence 1, Appli
35	79.5	4.5	1418	10	US-09-974-300-7280	Sequence 7280, Ap
c 36	79.5	4.5	2939	12	US-10-044-090-350	Sequence 350, App
37	79	4.5	1914	10	US-09-815-242-4167	Sequence 4167, Ap
38	79	4.5	1929	10	US-09-815-242-8045	Sequence 8045, Ap
39	78.5	4.4	1195	10	US-09-969-708-594	Sequence 594, App
40	78.5	4.4	1195	10	US-09-880-107-3746	Sequence 3746, Ap
41	78.5	4.4	2424	9	US-10-041-007-14	Sequence 14, Appl
42	78.5	4.4	2424	10	US-09-887-586A-45	Sequence 45, Appl
43	78.5	4.4	2424	10	US-09-903-012-45	Sequence 45, Appl
44	78.5	4.4	2528	9	US-10-041-007-13	Sequence 13, Appl
45	78.5	4.4	3731	10	US-09-822-268A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-848-852A-4
; Sequence 4, Application US/09848852A
; Patent No. US20020106373A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; Tang, Y. Tom
; Corley, Neil C.
; Guegler, Karl J.
; Yue, Henry
; Patterson, Chandra
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH EGF SIGNALING
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/848,852A
; FILING DATE: 07-May-2001

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/069,725
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0515 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1991 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BEPNOT01
; CLONE: 2056178
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-848-852A-4

Alignment Scores:
Pred. No.: 3,72e-221 Length: 1991
Score: 1772.00 Matches: 332
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-848-852A-3 (1-332) x US-09-848-852A-4 (1-1991)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHisLeuAla 20
DB 542 ATGTACTCTGCCAACCGTTTGGGGACAACTCTGGGCCAGAGGGGAGCGACATTGGCA 601
QY 21 ProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGln 40
DB 602 CCCCTGATGCCAGCATCCTCATCAGCAATGTGTGCAGCATCGGGGACCATGTGGCCAG 661
QY 41 GluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLys 60
DB 662 GAGCTTTTTCAGGGCTCAGATTTTGGGCTGCGCAGAGAGGCGAGAGCGCTGGGGAGAAA 721
QY 61 AlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAsp 80
DB 722 GCCGGCCAGACAGCCCCCTCGCAGAGAGAGCATGTGACCTCGCTACAGAGCATCTTGGAC 781
QY 81 GluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLys 100
DB 782 GAATTCCTTCAAAGGTATGGCAGCCTCATACCCCTCAGCATGATGAGGTAGTAGAGAAG 841
QY 101 LeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGln 120
DB 842 CTGGAGGACATTTTCCAGCAGGAGTTTTCACCCCTTCCAGGAAGGCGCTGGTTGCAG 901
QY 121 LeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAla 140
DB 902 CTGATCCAGCTCTACCAAGCGATGCCAGGCAATGCCATGTGTGAGGGCTTCCAGAGGGCT 961
QY 141 TyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeu 160
DB 962 TATAAGCGCAGCGTGTGACCATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCTC 1021
QY 161 AsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVal 180
DB 1022 AATGACAGGTGATGAACATGTATGGAGAGACCTGTGTCATGGACACAGATCCCTGAAAAAGGTG 1081
QY 181 HisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLys 200
DB 1082 CATTTCTCAATAGTTTCTTATGATATAAATCCGTACCAAGGGTATGATGGGGTAAA 1141
QY 201 ArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIlePheIleHisLeu 220
```

```
DB 1142 AGGTGGACCAAAAACGTGGACATCTTCAATAAGGAGGTACTGCTTAATCCCATCCACCTG 1201
QY 221 GluValHisTrpSerLeuIleSerValaspValargargthrIleThrTyrPheasp 240
DB 1202 GAGTGCATTGGTCCCTCATCTCTGTTGATGTGAGGGGACGACCATCACCATTATTTGAC 1261
QY 241 SerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGlu 260
DB 1262 TCGCAGCGTACCCTTAACCCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAGAG 1321
QY 261 AlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAsn 280
DB 1322 CGCGTAAAGAAAGACCGACTGGATTCCACGAGGCTGGAAAGGTTACTTCAAAATGAAT 1381
QY 281 ValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHis 300
DB 1382 GTGCCAGGCGAGATAATGACAGTACTGTGGTCTTTTGTGTTGCAGTACTGCAAGCAT 1441
QY 301 LeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGln 320
DB 1442 CTGGCCCTGTCTCAGCCATTTCAGCTTCACCCAGCAGACATGCCCAAACTTCGTCGCGAG 1501
QY 321 IleTyrLysGluLeuCysHisCysLysLeuThrVal 332
DB 1502 ATCTACAAGGAGCTGTGTCACTGCAAACTCAGCTGTG 1537
RESULT 2
US-09-848-852A-3 (1-332) x US-09-848-852A-4 (1-1991)
; Sequence 835, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 835
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 29-LIB188-017-Q1-E1-H1
US-09-848-852A-3 (1-332) x US-09-848-852A-4 (1-1991)
Alignment Scores:
Pred. No.: 8.19e-83 Length: 432
Score: 704.00 Matches: 135
Percent Similarity: 97.86% Conservative: 2
Best Local Similarity: 96.43% Mismatches: 3
Query Match: 39.73% Indels: 0
DB: 10 Gaps: 0
US-09-848-852A-3 (1-332) x US-09-848-852A-4 (1-1991)
QY 28 IleSerAsnValCysSerIleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAsp 47
DB 431 ATCAGCAACGTGTGCAGCATCGGGGACCATCGCCAGGAGCTTTTTCAGGGCTCAGAT 372
QY 48 LeuGlyMetAlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHisSerProLeu 67
DB 371 GTGGGCACTCAGAGAGGCTGAGAGGCTGGGGAGAAAGCTGGCCAGCAGCCCTCTG 312
QY 68 ArgGluGluHisValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGly 87
```

```

|||||
Db 311 CCGGAGGAGCATGTACCTGTGTGCAGAGCATCTTGGATGAATTCCTTCAGACTTACGGC 252
QY 88 SerLeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGln 107
Db 251 AGCCTCATACCCCTCAGACACTGATAGGTAGTGGAGAACTAGAGGACATTTCCAGCAG 192
QY 108 GluPheSerThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArg 127
Db 191 GAGTCTCTGCACCTTCCAGGAAGGCCCTGGTGTCTGCAGTTGATCCAGTCATACACGG 132
QY 128 MetProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThr 147
Db 131 ATGCAGGCAATGCCATGTGTAGGGGCTTCCGAGTCACCTATATAAGCGCATGTGTGACC 72
QY 148 MetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMet 167
Db 71 ATGGATGACCTGGGAACCTTTGTATGGACAACTGGCTCAATGACCAAGTGTGAACATG 12
RESULT 3
US-09-777-564-1710/c
; Sequence 1710, Application US/09777564
; Patent No. US2002022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mathion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 1710
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-777-564-1710

```

```

Alignment Scores:
Pred. No.: 3,7e-77 Length: 503
Score: 661.50 Matches: 125
Percent Similarity: 88.03% Conservative: 0
Best Local Similarity: 88.03% Mismatches: 0
Query Match: 37.33% Indels: 17
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x US-09-777-564-1710 (1-503)
QY 191 LeuArgThrLysGlyTyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsn 210
Db 502 CTCCTGACCAAGGGTATGATGGGTGAAAGGTGGACCAAAAAGCTGGACATCTTCAT 443
QY 211 LysGluLeuLeuLeuIlePheProIleHisLeuGluValHisTrpSerLeuIleSerValAsp 230
Db 442 AAGGAGCTACTGCTAATCCCATCCACCTGGAGGTGCATTTGGTCCCTCATCTCTCTGAT 383
QY 231 ValArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgCysPro 250
Db 382 GTGAGCGGACGCACCATCACCTATTATTGACTCGACGCTACCTAAACCGCGCTGCCCT 323
QY 251 LysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHis 270
Db 322 AGCATATATGCCAGTAATCTACAGCAGAGCGGGTAAAGAAACACCGACTGGATTTCCAC 263
QY 271 GlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAsnAspSerAspCys 290
Db 262 CAGGGCTGGAAGGTTACTTCAAAATG----- 236
QY 291 GlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThr 310
Db 235 -----TACTGCAAGCATCTGGCCCTGCTCTCAGCCCATTCAGCTTCACC 194

```

```

QY 311 GlnGlnAspMetProLysLeuArgArgGlnIleTyrLysGluLeuCysHisCysLysLeu 330
Db 193 CAGCAGGACATGCCCAACTTCGTGCGCAGATCTACAAGGAGCTGTGCTCACTGCAAACTC 134
QY 331 ThrVal 332
Db 133 ACTGTG 128
RESULT 4
US-09-856-247A-1
; Sequence 1, Application US/09856247A
; Patent No. US20020151028A1
; GENERAL INFORMATION:
; APPLICANT: Lima, Christopher
; APPLICANT: Mossesova, Elena
; TITLE OF INVENTION: Structure-Based Drug Design for Ulpi Protease Substrates
; FILE REFERENCE: 2650/LG681-US1
; CURRENT APPLICATION NUMBER: US/09/856,247A
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 60/205,336
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-856-247A-1
Alignment Scores:
Pred. No.: 6,21e-19 Length: 1866
Score: 222.50 Matches: 77
Percent Similarity: 43.46% Conservative: 56
Best Local Similarity: 25.16% Mismatches: 120
Query Match: 12.56% Indels: 53
DB: 10 Gaps: 12

US-09-848-852A-3 (1-332) x US-09-856-247A-1 (1-1866)
QY 36 AspHisValAlaGlnGluLeu---PheGlnGlySerAspLeuGlyMetAlaGluAla 54
Db 1063 GATTACTTAATAATCAAAAGTTGAAGTTTGATAGTATATATAGATTTGAGAAAGACTTC 1122
QY 55 GluArgProGlyGluLysAlaGlyGlnHisSerProLeuArgGluGluHisValThrCys 74
Db 1123 AAAAGATATACGAATTTTAATCAAAAGAGAGATTCAGAAGATCTTAAAAAAG 1182
QY 75 ValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThr 94
Db 1183 AAAGAACAATTCGCCAAGAAG-----AAACTTGTTCCT----- 1215
QY 95 AspGluValValGluLysLeuGluAspIlePheGlnGlnGluPheSerThrProSerArg 114
Db 1216 ---GAATTAATGAAAGACGATGACCAAGTACAAAAGCTTTGGCA-----TCTAGA 1266
QY 115 LysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetVal 134
Db 1267 GAAATACCTCACTTAATGNAATAGACATATATAGAG----- 1302
QY 135 ArgGlyPheArgValAlaTyrLysArgHisValLeuThrMetAspLeuGlyThrLeu 154
Db 1303 -----ATAACAGTACGTGATTTTAAAGACCTTG 1329
QY 155 TyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAsp 174
Db 1330 GCACCACGAAGATGCTTAATGACACTATCATTTAGTTTATGAAATFACATTGAAAAA 1389
QY 175 ThrValProGluLysValHisPheAsnSerPhePheTyrAspLysLeuArgThrLys 194
Db 1390 TCTACCCCTTAATACAGTG---CGGTTTATTCATTTTCTATACCAATTTATCAAAAG 1446
QY 195 GlyTyrAspGlyValLysArgTrpThrLys-----AsnValAspIlePheAsn 210
Db 195 ||||| |||||:||||| |||

```



```

; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 7902
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701100484H1
US-09-878-574-7902

Alignment Scores:
Pred. No.: 2,7e-08 Length: 264
Score: 130.50 Matches: 27
Percent Similarity: 52.50% Conservative: 15
Best Local Similarity: 33.75% Mismatches: 25
Query Match: 7.36% Indels: 13
DB: 10 Gaps: 2

US-09-848-852A-3 (1-332) x US-09-878-574-7902 (1-264)
QY 178 GluLysValHisPheAsnSerPheTyrAspLysLeu----- 191
Db 22 AAGAGGTTTCACCTTTTCAATTCCTTTTCTTCTTAAGCTGTTGATTAAAGCACAGAT 81
QY 192 ---ArgThrLysGlyTyrAspGly-----ValLysArgTrpThrLys 204
Db 82 AATCGTTCAATTGCTTGTGATGTAAGCAGCATTTTCAGCGTGTAAAGCAATTGGACAAGA 141
QY 205 AsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHisTrp 224
Db 142 AAGTGAACCTTTTGAAGAGGATTATATCTTCATTCCCATAAACTATAGTCTTCAATGG 201
QY 225 SerLeuIleSerValAspValArgArgTrpIleThrTyrPheAspSerGlnArgTrp 244
Db 202 AGTTGATGTCATTTGTCACCTGCTGAAGTCATGATGTCATGATGATGATGATGATGAT 261

RESULT 8
US-10-008-461-1
; Sequence 1, Application US/10008461
; Patent No. US20020127692A1
; GENERAL INFORMATION:
; APPLICANT: Barbara Ink
; TITLE OF INVENTION: New Protein
; FILE REFERENCE: QG1034(p80209)
; CURRENT APPLICATION NUMBER: US/10/008,461
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 0027905.9
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 639
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(636)
US-10-008-461-1

Alignment Scores:
Pred. No.: 1,25e-06 Length: 639
Score: 122.50 Matches: 48
Percent Similarity: 40.40% Conservative: 32
Best Local Similarity: 24.24% Mismatches: 73
Query Match: 6.91% Indels: 45
DB: 12 Gaps: 8

US-09-848-852A-3 (1-332) x US-10-008-461-1 (1-639)

QY 139 ValAlaTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlnAsn 158
Db 16 TTGAGTTATGACAGCTACTGTCGGCAATCAGATGTCCTACTATTGGATCGGCAAGC 75
```

```

QY 159 TrpLeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGlu 178
Db 76 TGGCTCAATGACCATATTATTGGTTTGGTTTGAG----- 111
QY 179 LysValHisPheAsnSerPhePheTyrAsp----- 189
Db 112 -----TACTTTGCCAACACAGTCAGTTTCATGACTGCTCTGATCACGTCAGTTTCATCAGC 165
QY 190 -----LysLeuArgThrLysGlyTyrAspGlyValLysArgTrp 202
Db 166 CCTGAAGTCACCCAGTTTCATCAAGTC-----ACTAGCAACCCAGCAGATTGCCATGTTCC 222
QY 203 ThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGlu--- 221
Db 223 CTTGAACCACTGGACCTCCCAACAGAGAGTGTATTTTAGCCATCAATGATAACTCC 282
QY 222 -----ValHisTrpSerLeuIleSerValAspValArgArgTrp 235
Db 283 AACAGGCAGCTGGAGGAACCCACTGGAGTTTATTGGTCTACCTCCAAGATAAAATAGC 342
QY 236 IleThrTyrPheAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLys 255
Db 343 TTTTTCATATGATTTCCCATAGCAGGCACTCAGTTCCACCAAGCAGGTAGCAGAG 402
QY 256 TyrLeuGlnAlaGluAlaValLysLys---AspArgLeuAspPheHisGlnGlyTrpLys 274
Db 403 AAACGGAGGCTTCTTAGCAGAAAGAGAGACAAACTGGCCTTT----- 447
QY 275 GlyTyrPheLysMetAsnValAlaArgGlnAsnAspSerAspCysGlyAlaPheVal 294
Db 448 -----GTGGAAGAGAAAGCCCTGCCCAACAAACAGCCTATGACTGTGGATGTACGTG 501
QY 295 LeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGln 312
Db 502 ATA-----TGAACACTGAGGCTTGTGTCAAGACTTC---TTTAGGCAACAG 546

RESULT 9
US-09-728-445-223
; Sequence 223, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
; FILE REFERENCE: Lex-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 223
; LENGTH: 100
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-728-445-223

Alignment Scores:
Pred. No.: 5,11e-07 Length: 100
Score: 115.50 Matches: 22
Percent Similarity: 84.85% Conservative: 6
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 6.52% Indels: 1
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x US-09-728-445-223 (1-100)

QY 180 ValHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVal 199
Db 1 GTCCACTTCTTCAACAGCTTTTCCATAGACAGCTGGTAACCAAGGCTATAATGAGT 60
```


Db	463	CCAGATTCGTCGTTATTTCCTTTTCAACCTTTACCGAAT-----	498
Qy	42	LeuPheGlnGlySerAspLeuGlyMetAlaGluGluArgProGlyGluLysAla	61
Db	499	TTGTATCCTGTCACACCTTTAGCAACT-----GATATAGTT	534
Qy	62	GlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAspGlu	81
Db	535	GGAAGTGAGGAGTCATTTCCCAAATCAATCTAACTAATTTGCAGAAAAATTTACAAAG	594
Qy	82	PheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValValGluLysLeu	101
Db	595	TTTTTCAAAACCTGTAACATGCTCTCGTTTTTACTTGGTAAATTTTCATGTGGACCGAGTA	654
Qy	102	GluAspIlePheGln-----GlnGluPheSerThrProSerArgLys	115
Db	655	CAGGACTATTTTGAAGCAAAAGAACTGAAAGATTCAGATTTTCAGGAAGTAGCAAGAGAA	714
Qy	116	GlyLeuValLeuGlnLeuIleGlnSerTyrglnArgMet-----Pro	129
Db	715	AAGTGTGTTTTTACAGCCTGTAAAGCCACAGATAGTAGAATGGAAGTATCTCTCC	774
Qy	130	GlyAsnAlaMet--ValArgGlyPheArgVal-----AlaTyrlLysArg	143
Db	775	AAACTACGGATGGAGTTAGAGGTACCGAGAGAGTTCTTCAACCGGATGCTATCGACAT	834
Qy	144	HisValLeuThrMetAspAspLeuGlyThrLeuTyrglyGlnAsnTrpLeuAsnAspGln	163
Db	835	CATATTTTATAAAATATTGTTTGGCAATGATCTTTGGT-----TGGACTTCGGATCGT	888
Qy	164	ValMetAsnMetTyrglyAspLeuValMetAspThr-----Val	176
Db	889	TTTCAAAATGTTATGAATCAGGTAAATGATGCATCCCTTATCTCTGGAAGTTGAAGTA	948
Qy	177	ProGluLysValHisPhePheAsnSerPhePheTyrlAspLys-----LeuArg	192
Db	949	ACAAGTCGCTTCATTTGTCTGATTCACAATGGATACGAAAGCCAGTTCCTTTGTCT	1000
Qy	193	ThrLysGlyTyrlAspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGlu	212
Db	1009	CATCAATTTAGGAAGGCTATTCGTAAATTTTCAAAAGATTAGATATTCACAGGAACAT	1066
Qy	213	LeuLeuLeuIleProIleHisLeu-----GluValHisTrpSerLeuIleSerValAsp	230
Db	1069	TTAGATATTATCAAAAGAGAGATGTTGGCGAATTTTTCAGTAGCATGAACCTCTCTTGA	112
Qy	231	ValArgArgArgThrIleThrTyrlPheAspSerGlnArgThrLeuAsnArgArgCysPro	250
Db	1129	TTTTTATCCACCGCAATATGATGCTTTTGAAATGGTCAGACAAATT-----	117
Qy	251	LysHisIleAlaLysTyrlLeuGlnAlaGluValLysLysAspArgLeuAsp	268
Db	1174	TTTCATTTGCGGAAAAATTTTACAG--GAAATTTACTTTAGAGATGTCCTTGAAT	1224

Alignment Scores:	
Pred. No.:	0.0578
Score:	90.50
Percent Similarity:	29.21%
Best Local Similarity:	20.86%
Query Match:	5.11%
DB:	10
Length:	1384
Matches:	58
Conservative:	51
Mismatches:	11
Indels:	55
Gaps:	12

QY 22 ProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGlnGlu 41

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 19:36:15 ; Search time 257 seconds
(without alignments)
124.189 Million cell updates/sec

Title: US-09-848-852A-3
Perfect score: 1772
Sequence: 1 MYSAQREWGTIWARRGAHLA.....DMPKLRQIYKELCHCKLTV 332

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1667	94.1	371	2 T08759	conserved hypothe
2	319.5	18.0	568	2 T39840	conserved hypothe
3	268.5	15.2	662	2 T16845	hypotheical prote
4	240.5	13.6	233	2 T01233	hypotheical prote
5	222.5	12.6	621	2 S63462	hypotheical prote
6	205.5	11.6	652	2 T37822	probable centromer
7	197.5	11.1	424	2 H85175	hypotheical prote
8	196	11.1	1034	2 S49947	SMT4 protein - yea
9	193	10.9	547	2 T02274	hypotheical prote
10	190	10.7	398	2 E71424	hypotheical prote
11	163	9.2	710	2 T04979	hypotheical prote
12	156	8.8	1240	2 T48800	SMT4 related prote
13	136	7.7	920	2 T36650	hypotheical prote
14	124.5	7.0	342	2 T19877	hypotheical prote
15	118.5	6.7	311	2 T32915	hypotheical prote
16	116.5	6.6	662	2 T33900	hypotheical prote
17	109.5	6.2	1121	2 T13796	kinesin-related pr
18	109	6.2	1305	2 H96559	hypotheical prote
19	107.5	6.1	272	2 C86642	transposase of IS1
20	106	6.0	418	2 D71460	probable membrane
21	105	5.9	1153	2 F84468	hypotheical prote
22	105	5.9	1272	2 C64513	hypotheical prote
23	104.5	5.9	272	2 C86884	transposase of IS1
24	103	5.8	1218	2 E84537	hypotheical prote
25	102	5.8	550	2 B84500	hypotheical prote
26	101.5	5.7	780	2 B84084	ATP-dependent DNA
27	99.5	5.6	1121	2 T13750	inesin-like prote
28	98	5.5	227	2 D86631	transposase of IS1
29	98	5.5	1314	2 T47331	hypotheical prote

30	97	5.5	625	2 G72420	oligopeptide ABC t
31	97	5.5	790	1 WMFM9	90K protein - alfa
32	96.5	5.4	272	2 D86671	transposase of IS1
33	95.5	5.4	543	2 T02022	hypotheical prote
34	95.5	5.4	1201	2 F86386	hypotheical prote
35	94.5	5.3	929	2 B81982	probable isoleucin
36	94	5.3	227	2 G86704	transposase of IS1
37	93.5	5.3	662	2 E97738	DNA topoisomerase
38	93.5	5.3	1472	2 A84470	hypotheical prote
39	93	5.2	445	2 F83881	hypotheical prote
40	92.5	5.2	273	2 A42549	hypotheical prote
41	92	5.2	1198	2 E86402	hypotheical prote
42	91.5	5.2	353	2 D81273	probable 3-oxoacyl
43	91.5	5.2	661	2 T30179	hypotheical prote
44	91	5.1	1175	2 D85089	hypotheical prote
45	90.5	5.1	297	2 C70231	hypotheical prote

ALIGNMENTS

RESULT 1

T08759
conserved hypothetical protein DKFp586K0919.1 - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
C:Accession: T08759
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z16471
A:Accession: T08759
A:Molecule type: mRNA
A:Residues: 1-371 <WAM>
A:Cross-references: EMBL:AL050283
A:Experimental source: adult uterus; clone DKFp586K0919
C:Genetics:
A:Note: DKFp586K0919.1

Query Match 94.1%; Score 1667; DB 2; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.5e-130; Mismatches 0; Indels 0; Gaps 0;
Matches 314; Conservative 0;

Qy	19	LAPPDASILISNVCISGIDHVAQELFGQSDLGMAEEAERPEKAGQHSPLREHRTVCQSI	78
Db	58	LAPPDASILISNVCISGIDHVAQELFGQSDLGMAEEAERPEKAGQHSPLREHRTVCQSI	117
Qy	79	LDEFLTQYGSLLPLSTDEVVEKLEDFQOEFTSPSRKGLVLQLIQSYQRMPCNVMVGRPR	138
Db	118	LDEFLTQYGSLLPLSTDEVVEKLEDFQOEFTSPSRKGLVLQLIQSYQRMPCNVMVGRPR	177
Qy	139	VAYKRHVLTMDDLGTLYGONLNDQVMNMGDLVMDTVPEKVHFFNSFFYDKLRTKGYDG	198
Db	178	VAYKRHVLTMDDLGTLYGONLNDQVMNMGDLVMDTVPEKVHFFNSFFYDKLRTKGYDG	237
Qy	199	VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRCRPHIAKYIQ	258
Db	238	VKRWTKNVDIFNKELLIPILHLEVHWSLISVDVRRRTITYFDSQRTLNRCRPHIAKYIQ	297
Qy	259	AEAVKRDRLDFHQWKGKFKMNVARQNNDSCGAFVLQYCKHLALSQPFSTQDMPKLR	318
Db	298	AEAVKRDRLDFHQWKGKFKMNVARQNNDSCGAFVLQYCKHLALSQPFSTQDMPKLR	357
Qy	319	ROIYKELCHCKLTV	332
Db	358	ROIYKELCHCKLTV	371

RESULT 2

T39840
conserved hypothetical protein SPBC19G7.09 - fission yeast (Schizosaccharomyces pombe
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39840

R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z21884
A:Accession: T39840
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-568 <LN>
A:Cross-references: EMBL:AL021839; PIDN:CAA17063.1; GSPDB:GNO0067; SPDB:SPBC19G7.09
A:Experimental source: strain 972h-; cosmid c19G7
C:Genetics:
A:Gene: SPDB:SPBC19G7.09
A:Map position: 2

Query Match 18.0%; Score 319.5; DB 2; Length 568;
Best Local Similarity 28.7%; Pred. No. 1.6e-18;
Matches 90; Conservative 51; Mismatches 114; Indels 59; Gaps 9;
QY 59 EKAGQHSPLEEHVTCVQS-----ILDEFLQYGSILPLSTDEVEKLEDFIQ----- 107
DB 264 KSPHDSTSRKASPRFYQSDQPARNIIVTSDIQNEKSLLLIIRD-LKEKQTESFQDQWNEV 322
QY 108 -----EFSPSRKGLVLQLIQSYQVRMPGNAMVGRFVAKRH 144
DB 323 DFQLKGLSTPPTRPKFPELEFPDNRKALKYLQNSVSSSPIITKFNIP----- 378
QY 145 VLTMDLGLTYGQNLNDQVMNMYGDLV-----MDTVPEKVVHFFNSFFYDKLRTKGYDGV 199
DB 379 -ITLKLHLTLNRQWLNDENVYFMNLISERSKIDSSLPRVHGNTFFYTSLQRRGAGV 437
QY 200 KRWTG--NVDIFNKELLIPIHLEVHWSLISVDVRRRTTYFDSQRTPLNRRCPKHIAYL 257
DB 438 RRAKARVNIADMDAVFIPVHLVDHMCMAVINKSKRFEYWDLSLSPGKGVFDLLRDY 497
QY 258 QAEAVKDRDLDFHQGWKGYFKMVARQNDSCGAFVLQYCKHLALSQPSFTQDMPKL 317
DB 498 IAET--KGAVDV--SDWENFMDNSPQRNGHDCGVFACTAECVSRVNPVQFSQNDMPEL 554
QY 318 R-----RQIV 322
DB 555 RIKMAASIIDAQIV 568

RESULT 3
T16845
hypothetical protein T10F2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16845
submitted to the EMBL Data Library, March 1995
A:Description: The sequence of C. elegans cosmid T10F2.
A:Reference number: Z18589
A:Accession: T16845
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-662 <MIL>
A:Cross-references: EMBL:U234412; NID:g727446; PID:g727449; PIDN:AAA64293.1; CESP:T10F2.3
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:T10F2.3
A:Introns: 17/1; 51/1; 96/3; 197/3; 291/3; 342/3; 375/2; 619/2; 645/3; 649/1

Query Match 15.2%; Score 268.5; DB 2; Length 662;
Best Local Similarity 26.4%; Pred. No. 3.3e-14;
Matches 75; Conservative 53; Mismatches 79; Indels 77; Gaps 10;
QY 27 LISNVCISG---DHVAQELFQGSDLGMAEAPRGKAGQHSPLREHVTVCQSILDEF 83
DB 396 ILSQINSIGIGSAYRGPRQYONS---YQLSKQKEDKLLLEEARIREGHS----- 441
QY 84 QTYGSLIPLSTDEVEKLEDFIQEFTSPSRKGLVLQLIQSYQRM----- 128
DB 442 QTRG-----DRLEDV-----RKRLQGIATIRPKVEKKKVDVDFMALPDAA 481

QY 129 -----PGNAMVGRFVAYKRRLVLTMDLGLTYGQNLNDQVMNMYGDLVM--- 173
DB 482 DALVERAWSGGNPEQVDAFSIQ-----ICKKDLATLSGLHLWLNDEIINFYLQLICDRS 536
QY 174 --DTVPEKVVHFFNSFFYDKLRTKGYDGVKRWTKNVDIFNKELLIPIHLEVHWSLISVDV 231
DB 537 NGDSKYPKIYAFNTFFYSNIVSGYASVKKRWTRKVDIFAFDVLVPHLGMHMCMAVIDM 596
QY 232 RRTTYFDSQRTLNRRCPKHIAYLQAEAV--KKDRLDPHQGW 273
DB 597 GERKIEFYDSLDYDNTAVLPLPALRGYLEAESLDRKKKTAMNF-SGW 639

RESULT 4
T01233
hypothetical protein F6N23.7 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C:Accession: T01233
R:Geisel, C.
submitted to the EMBL Data Library, April 1998
A:Description: The sequence of A. thaliana F6N23.
A:Reference number: Z14281
A:Accession: T01233
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-233 <CEI>
A:Cross-references: EMBL:AF058919; NID:g3047100; PID:g3047118; GSPDB:GNO0063; ATSP:F6
C:Genetics:
A:Gene: ATSP:F6N23.7
A:Map position: 5
A:Introns: 9/1; 21/2; 52/3; 85/3; 115/3; 147/3; 182/2; 213/3

Query Match 13.6%; Score 240.5; DB 2; Length 233;
Best Local Similarity 31.3%; Pred. No. 1.7e-12;
Matches 62; Conservative 32; Mismatches 75; Indels 29; Gaps 7;
QY 149 DDLGLTYGQNLNDQVMNMYGDLVMDTV---PE---KVHFFNSFFYDKLRT---KGYDGV 199
DB 38 ETQLCLRPQWLNDVTLNLYLELLKERQTRDPQKYFKCHFFNTFFYKLVSGSGYNYKAV 97
QY 200 KRWTG----NVDIFNKELLIPIHLEVHWSLISVDVRRRTTYFDSQRTLNRRCPKHIAK 255
DB 98 SWTTTKRKLGYDLIDCDIIFVPIHIDIHTLGVINNRRKRVYLDLSFT-----GAK 149
QY 256 YLQAEAVKDRDLDFH-QGWKGYFKMVARQNDSCGAFVLQYCKHLALSQPSFTQ--- 311
DB 150 YLVDEVKQSKQKNIDVSSWGMVEYVEERPOQNGYDCGMFLKYIDFYSRGLSLQFSQVIR 209
QY 312 ----QDMPKLRRQIYKEL 325
DB 210 DVTKDMPYFRJRTAKEI 227

RESULT 5
S63462
hypothetical protein YP020c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein LPB1lc
C:Species: Saccharomyces cerevisiae
C:Date: 16-May-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
C:Accession: S63462
R:Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.;
submitted to the EMBL Data Library, September 1995
A:Reference number: S63452
A:Accession: S63462
A:Molecule type: DNA
A:Residues: 1-621 <WAN>
A:Cross-references: EMBL:U36624; NID:g1276642; PID:g1039457; GSPDB:GNO0016; MIPS:YPL0
C:Genetics:
A:Gene: SGD:ULP1; MIPS:YPL020c
A:Cross-references: SGD:S0005941
A:Map position: 16L


```

Query Match      12.6%; Score 222.5; DB 2; Length 621;
Best Local Similarity 25.2%; Pred. No. 2e-10;
Matches 77; Conservative 56; Mismatches 120; Indels 53; Gaps 12;

QY 36 DHVAOEL-FQSGDLGMAEEAERPECEKAGQHSPLREEHVTCVQSILDFELOYGSLIPLST 94
Db 355 DYLNQKLFORSILFEKDFKARYNEILNERKKIIOELDKKKEQJAKK-----KLVP--- 405
QY 95 DEVVEKLEDFIQOBFSTPSRKGLVLQLOTSQYORMPGNAVGRFVAVKRRHVLTMDDLGTL 154
Db 406 -ELNEKDDQVQKALA--SRENTOLMNRDNI-----ITVRDFKTL 443
QY 155 YGQWLNDQVNMVYGDVLVMDTVPEKVFHFFNSFFYDKLTKGYDGVKRWTK----NVDIFN 210
Db 444 APRRWLNDTIEFPKVIKSTPNTV-AFNSFFYTNLSERCYQGVRRMMKKTKQIDKLD 502
QY 211 KELLIIPIHL-EVHWSLISVDVRRRTITYFDS----QRTLNRRCPKHIKYLQAEAVKKD 265
Db 503 K-IFTPINLQNSHWALGIIDLKKKTTGYVDSLSGPNAMSFALITDLQKYVMEESKHTI 560
QY 266 RLDPFHQGWKGYFKMNAVRQNNDSDCGAFVLQYCKHLLASQPSFTQODMPKLRRIQYKEL 325
Db 561 GEDF-----DLIHLDCQQPNGYDCGIYVCNMTLYGSADAPLDFDYDAIRRRRI----- 611
QY 326 CHECKLT 331
Db 612 AHLILT 617

RESULT 6
T37822
probable centromere protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T37822
R;Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z21737
A;Accession: T37822
A;Status: preliminary; translated from GB/EMBL/DDBB
A;Molecule type: DNA
A;Residues: 1-652 <DEV>
A;Cross-references: EMBL:Z98489; PIDN:CAB11507.1; GSPDB:GN00066; SPDB:SPAC17A5
A;Experimental source: strain 972h-; cosmid c17A5
C;Genetics:
A;Gene: SPDB:SPAC17A5.07c
A;Map position: 1
A;Introns: 2/2; 105/1; 387/2; 422/2; 501/2

Query Match      11.6%; Score 205.5; DB 2; Length 652;
Best Local Similarity 26.0%; Pred. No. 5.5e-09;
Matches 60; Conservative 36; Mismatches 72; Indels 63; Gaps 7;

QY 146 LTMDDLGTLGQGNWLNQVNMVY-----GDLVMDTVPEKVFHFFNSFFYDKLRT 193
Db 363 ITNTDLRLNEGEFLNDITVDFYLRVLYCKLQTNQPSLANDT----HIFWTFYNRLTS 417
QY 194 K-----GVDGVRKWTKNVDIFNKELLPIHLEVIHV-----SLISVDVRR--- 234
Db 418 KDKDGKRLGHRGVKKTQKVDLFHKYIIVINETFHWYLAIIICNIDRLMPDVTKLEBQD 477
QY 235 -----TITYFDSQRTLNRRCPKHIKYLQAEAVKKDRDLF 269
Db 478 EIVMSSVEQPSASTROAELTNSPALIFDPSLANLHKGALNYLREYLLEAEAFERKNVHL 537
QY 270 HGWKGYFKMNAVRQNNDSDCGAFVLQYCKHLLASQPSFTQODMPK-LRR 319
Db 538 KSTDIRGFHAKVPOQSNFSDCGIYALHFEV-LFLETPEQVIANTIDKSLRR 587

RESULT 7
H85175

```

hypothetical protein ATg1580 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory

Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: AB5001; MUID:2008348; PMID:10617198
A:Accession: H85175
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-424 <STO>
A:Cross-references: GB:NC_001268; NID:g7268336; PIDN:CAB78630.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g15880
A:Map position: 4

Query Match 11.1%; Score 197.5; DB 2; Length 424;
Best Local Similarity 23.8%; Pred. No. 1.4e-08;
Matches 74; Conservative 56; Mismatches 114; Indels 67; Gaps 12;

QY 9 GTIWARGAHLAPPDASTLIISNVCSIGD-----HVAQELFGSDDLGMAEAEERPGEK 60
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 125 GVTVDKGENFRVEDTSMWLSLSDRDNDASSLEAYRKLMQS-----REKRNSKLEA 179
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 61 AGQHSPLEAREHVTCVQSILDFLOTYGSLPLSTDVEVVEKLEDIFQQEFTSPSRKGLVLQ 120
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 180 LGFEIVLNKKL-----SLLQRSRPKTVEKRVVPREPFP-----IP 215
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 121 LIQSYQRMPGNAVRGRVAVKRHVLF-----TWDDLTGYQGWNLDVNNMVGTD 170
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 216 LTEDEA---EVYRAFSGRRNRKVLAATHENSIDITGEVLQCLTPSAWLNDVINLVLE 271
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 171 LVMDTVPE-----KVHFENSEFFDYDKL-RTGYD-GVKRWTK----NVDLPNKELLIP 217
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 272 LKKEETREPKYLKCHFVNTFFFKLVDSGSYNFKA VRWTQTKLGVALIDCDMI FVP 331
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 218 IHLEVHWSLISDVYRRRTITYFDSORTLNRRCPRKHIAKYLQA EAVKGD--RLDPHQGWKG 275
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 332 IHRGVHTLAINNRRESKLLYLDLSINGVDPMILALAKYMGD EANEKSKKIDAN-SWDM 390
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 276 YFKNAVRAQN 286
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 391 EFVEDLPQQKN 401
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 8
S49947
SMT4 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein Y19905.17; protein Y1L031w
C:Species: Saccharomyces cerevisiae
C>Date: 13-Jan-1995 #sequence_revision 20-Feb-1995 #text_change 06-Feb-1998
C:Accession: S49947; S59668
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, December 1994
A:Reference number: S49931
A:Accession: S49947
A:Molecule type: DNA
A:Residues: 1-1034 <ODE>
A:Cross-references: GB:I47047; EMBL:I46861; NID:g603997; PID:g7633315; MIPS:Y1L031w
R:Maluh, P.B.; Koshland, D.E.
submitted to the EMBL Data Library, May 1995
A:Reference number: S59668
A:Accession: S59668
A:Molecule type: DNA
A:Residues: 1-493, 'XX', 496-870, 'X', 872-1034 <MEU>
A:Cross-references: EMBL:U27832; NID:g886765; PID:g886766
C:Genetics:
A:Gene: SGD:SMT4
A:Cross-references: SGD:S0001293; MIPS:Y1L031w
A:Map position: 9L

Query Match 11.1%; Score 196; DB 2; Length 1034;

Best Local Similarity 27.2%; Pred. No. 6.2e-08;
Matches 52; Conservative 33; Mismatches 64; Indels 42; Gaps 5;

QY 144 HVLTMDDLGLTYGONWLDQVMNMYGDLVMDTVPEK-----VHFFNSFFYDKLRPKG- 195
Db 442 YTTNQDFKCLFNKQWNSLDLFFTKFYFESSIEKSIIRKQVHLMSSFFYTKLISNPA 501

QY 196 --YDGVRRKTKNVDIFNKELLIPILHLEHVH-----SLI 227
Db 502 DYISNVKKWNTDLFSKKYVVPINISYHWFSCIITNLDAILDHONKDKNDAINSDEI 561

QY 228 SDVVRRTITYFDSQRTLNRCRPHIAKYLOAEAVKKDRDLDFHQGWGYFKMN---VARQ 284
Db 562 SINPLNVLITFDSLRQTHREIDPIKEFLISYALDKYSIQLD---KTQIKMKTKCPVPQ 618

QY 285 NNDSDCGAFVL 295
Db 619 PNMSDCGVHVI 629

RESULT 9
T02274
hypothetical protein T13D8.11 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 22-Oct-1999
R:Vysotskaia, V.S.; Schwartz, J.R.; Kwan, A.; Toriumi, M.; Yu, G.; Oji, O.; Liu, S.; Li-
rtz, D.; Li, Y.; Palm, C.J.; Shim, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N.
submitted to the EMBL Data Library, June 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC T13D8 sequence.
A:Reference number: Z14649
A:Accession: T02274
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-547 <Y>
A:Cross-references: EMBL:AC004473; NID:g3108025; PID:g3249071; GSPDB:GN00059; ATSP:T13D8
C:Genetics:
A:Gene: ATSP:T13D8.11
A:Map position: 1
A:Introns: 106/1; 154/3; 259/2; 275/3; 304/2; 332/2; 366/3; 399/2; 440/2; 478/3; 514/3

Query Match 10.9%; Score 193; DB 2; Length 547;
Best Local Similarity 25.9%; Pred. No. 4.7e-08;
Matches 67; Conservative 43; Mismatches 85; Indels 64; Gaps 12;

QY 92 LSTDEVVEKLEDFQOEFSPSRKGLV-LQLIQSYQRMPCGNAMVGRVAYKRHLVTMD 150
Db 282 INVDEEAQSPVAEQAEUP--EGLIKLQL-----ATYK---LIVDK 319

QY 151 LGTL-----YQCNWLDQVMNMYGDLVMDTVPEKVHFFNSFFYDKLRT-----KGYD---- 197
Db 320 TCSLQEDICYPTRFLOQIIS-----SNOISADCHFFNTFYFKKLSDAVYKGNKDQAF 373

QY 198 --GVKRWTKNVDIFNKELLIPILHLEHVHSLISV-----DVRRRTITYFDSQRTLNRR- 248
Db 374 FVFRFRWKKGIDLFKAYIPIPIHEDLHWSLVTVICIPDKRDESLTILHDSLGLHSRKS 433

QY 249 CPKHIKYLQAEAVKKDRDLDF-----HGWKGYEK-----MNVARQNNDSDCGAFVL 295
Db 434 IVENVRFLKADENYLNQDDYSLDLPISKVKWNLPRRISEAVVQVPQKNQDFDCGPFVL 493

QY 296 QYCKHLALSQFFSFTQODM 314
Db 494 FFIKRTIEEAPQRLKRKDL 512

RESULT 10
E71424
hypothetical protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 11-Jan-2002
C:Accession: E71424

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; G
avanagh, T.; Hempel, S.; Kottler, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: E71424
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-398 <BEV>
A:Cross-references: GB:Z97339; NID:g2244901; PID:g2244945
C:Genetics:
A:Map position: 4COP9-4G3845

Query Match 10.7%; Score 190; DB 2; Length 398;
Best Local Similarity 24.0%; Pred. No. 5.5e-08;
Matches 69; Conservative 50; Mismatches 104; Indels 64; Gaps 10;

QY 9 GTIWARGAHLAPPDASILISNVCISGD-----HVAQELPQSGDLGMAEEAERPGEK 60
Db 125 GYTDVAKKGNFRYEDTSMMLDSLSLDRVDNDASSLEAYRKLMQS-----AEKRNKLEA 179

QY 61 AGOHSPLREEHVTVCQVQSLDEFLQTYGSLIPSTDEWKELEDFQOEFSTPSRKGLVLQ 120
Db 180 LGFEIVLNEKKL-----SLLQSRPKTVKEKRVVPREPF-----IP 215

QY 121 LTIQSYQRMPCGNAMVGRVAYKRHLV-----TMDLGLTYGONWLDQVMNMYGD 170
Db 216 LTEDDEA-----EYVRAFSGRNRKRLVATHENSINIDITGEVLQCLTPSAWLNDVINVYLE 271

QY 171 LVMDTVPE-----KVHFFNSFFYDKL-RTKGYD--GVKRWTK-----NVDIFNKELLIP 217
Db 272 LKERETREPKYKLYCHYNTFFYKLVSDSGYNFKAVRVFTQKRLGVALIDCDMIFVP 331

QY 218 IHLEVHWSLISVDVRRRTITYFDSQRTLNRCRPHIAKYLOAEAVKK 264
Db 332 IHGCVHTLAVNNRESKLLYLDLSLNGVDPMILNALAKYMGDEANEK 378

RESULT 11
T04979
hypothetical protein T16L1.110 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C:Accession: T04979
R:Bevan, M.; Obermaier, B.; Deutschenbaur, S.; Piravandi, E.; Hoheisel, J.; Mewes, H.
submitted to the Protein Sequence Database, November 1998
A:Reference number: Z15393
A:Accession: T04979
A:Molecule type: DNA
A:Residues: 1-710 <BEV>
A:Cross-references: EMBL:AL031394
A:Experimental source: cultivar Columbia; BAC clone T16L1
C:Genetics:
A:Map position: 4
A:Introns: 21/3; 97/3; 123/1; 141/1; 212/2; 279/2; 298/1; 327/2; 407/2; 518/2; 626/2
A:Note: T16L1.110

Query Match 9.2%; Score 163; DB 2; Length 710;
Best Local Similarity 31.7%; Pred. No. 2.1e-05;
Matches 32; Conservative 21; Mismatches 32; Indels 16; Gaps 2;

QY 145 VLTMDLGLTYGONWLDQVMNMYGDLVMDTVPEK-----VHFFNSFFYDKLRT----- 193
Db 187 VVRKQDIELLKPRRFINDTIIDFYIKYLNRIKSPKGRGRHFHFNCFRKLANKDGTGS 246

QY 194 -----KGYDGVKRWTKNVDIFNKELLIPILHLEHVHSLISV 229
Db 247 TCGGREAYQRVKWTKNVDLFKDYIFIPINCSFHWSLVII 287

Db 247 AQARTV-IFDSQADLNQNMATLI--ESFMKYSYEKRTGNAMPFPLPCILPQRMPOQ 303
QY 285 NNSDCGAFVLQYCKHLALSOP 306
Db 304 TNFDCGIFIAEFARFLLSPP 325

RESULT 15

T32915
hypothetical protein K02F2.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000

C:Accession: T32915
R:Maggi, L.; Goela, D.

submitted to the EMBL Data Library, January 1998

A:Description: The sequence of C. elegans cosmid K02F2.

A:Reference number: Z21246

A:Accession: T32915

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-311 <MAG>

A:Cross-references: EMBL:AF043699; PIDN:AAB97567.1; GSPDB:GN00019; CESP:K02F2.4

A:Experimental source: strain Bristol N2; clone K02F2

C:Genetics:

A:Gene: CESP:K02F2.4

A:Map position: 1

A:Introns: 16/3; 72/2; 103/3; 139/2; 203/2; 238/2; 273/3

C:Superfamily: Caenorhabditis elegans hypothetical protein K02F2.4

Query Match 6.7%; Score 118.5; DB 2; Length 311;
Best Local Similarity 19.8%; Pred. No. 0.034;
Matches 51; Conservative 35; Mismatches 80; Indels 91; Gaps 10;

QY 142 KRHLTMDDDLGTLGQNW-----LNDQVMNYG-----DLVMDTVPEKVHFFNSF 186

Db 28 RRHIIP-----PLFHNGWVGVRNEDRTLLNDTIIEFYMCWMRLEVFDEATRASSHVHFSF 82

QY 187 FYDKLRT-----KGYDGVKRWTKNVDIF-NKELLIPIHLE--VHWSLI 227

Db 83 FLPIKTCFKDFENPPRSELANHYNRRFRSKNDATFLKDKILLIPVHLDKPKHWFLV 142

QY 228 SV-----DVRRTITYFDSQRTLNRR-----CPK----- 251

Db 143 IVHNPAGVRRISDVNILDATNKVSRRLSKRITGHVNCNENAGECRIIIMDSIVHSKY 202

QY 252 -----HIAYLOAEAVKKDRDLDFHQGWKGYFKMNVARQNDSDCGAFVLQYC 298

Db 203 REVIDKTHDSTFDHRLWLLMSAAATDVMFCTFRKVKVCQKLPQQRKNSVDCGIFMMAFA 262

QY 299 KHLALSQPSFTQODMP 315

Db 263 EYFT---KYNTAQSLP 276

Search completed: December 16, 2002, 20:15:46

Job time : 259 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2002, 19:28:45 ; Search time 207 Seconds
(without alignments)
66.522 Million cell updates/sec

Title: US-09-848-852A-3
Perfect score: 1772
Sequence: 1 MYSQREWGTTIWARRAHGLA.....DMPKLRQIYKELCHCKLTV 332

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	319.5	18.0	568	ULP1_SCHPO	Q42957: Schizosaccharomyces pombe (Fission yeast).
2	268.5	15.2	659	YRM3_CAEEL	Q09353 caenorhabdi
3	222.5	12.6	621	ULP1_YEAST	Q02724 saccharomyc
4	205.5	11.6	652	ULP2_SCHPO	Q13769 schizosacch
5	196	11.1	1034	ULP2_YEAST	P40537 saccharomyc
6	170.5	9.6	984	SUS2_HUMAN	Q09qf6 homo sapien
7	143	8.1	1112	SUS1_HUMAN	Q09qz1 homo sapien
8	124.5	7.0	342	YOG6_CAEEL	Q09275 caenorhabdi
9	105	5.9	1272	Y28_METUA	Q60287 methanococc
10	97	5.5	790	V90K_AWVLE	P03593 alfalfa mos
11	92.5	5.2	273	VPR1_ASEB7	Q00946 african swi
12	90	5.1	727	YR22_CAEEL	Q09639 caenorhabdi
13	89	5.0	608	HTPG_CAMEJ	Q09phz3 campylobact
14	87.5	4.9	273	VPR1_ASEF2	Q65228 african swi
15	87.5	4.9	694	DNJ1_MYCLE	O33102 mycobacteri
16	87.5	4.9	2034	FER1_CAEEL	Q17388 caenorhabdi
17	87.5	4.9	4344	DVHC_EMENI	P45444 emericeella
18	86.5	4.9	382	HUGA_APIME	Q08169 apis mellif
19	86.5	4.9	484	R54_BRAJA	P30332 bradyrhizob
20	86	4.8	824	TG37_HUMAN	Q13099 homo sapien
21	85.5	4.8	538	Z155_HUMAN	Q12901 homo sapien
22	85	4.8	300	HTPG_MOUSE	Q09wt0 m geranylge
23	85	4.8	607	HTPG_FUSNN	Q08gh4 fusobacteri
24	85	4.8	735	RN5A_MOUSE	Q05921 mus musculu
25	85	4.8	802	PAC_ARTV1	P31956 arthrobacte
26	85	4.8	864	RA50_SULSO	Q097wh0 sulfolobus
27	84.5	4.8	562	SIR2_YEAST	P06700 saccharomyc
28	84.5	4.8	626	HTPG_BACSU	P46208 bacillus su
29	84.5	4.8	653	YJU4_YEAST	P39531 saccharomyc
30	84	4.7	402	ELAD_ECOLI	Q47013 escherichia
31	84	4.7	476	MPBP_NEUCR	P11913 neurospora
32	84	4.7	731	NGP1_HUMAN	Q13823 homo sapien
33	83.5	4.7	266	Y209_METUA	Q57662 methanococc

34	83.5	4.7	294	1	GGPP_BOVIN	P56966 b geranylge
35	83.5	4.7	4725	1	DYHC_DICDI	P34036 dictyosteli
36	83	4.7	366	1	MT52_METJA	Q58600 methanococc
37	83	4.7	503	1	FUTA_DROME	Q9vul9 drosophila
38	83	4.7	2386	1	RAD3_SCHPO	Q02099 schizosacch
39	82.5	4.7	420	1	YFJD_HAEIN	Q07017 haemophilus
40	82.5	4.7	459	1	YLD8_STRCO	O86835 streptomyce
41	82.5	4.7	625	1	RSC4_YEAST	Q02206 saccharomyc
42	82.5	4.7	1587	1	TOP2_PENCH	Q9y8g8 penicillium
43	82	4.6	353	1	LICH_PSEGL	Q05490 pseudomonas
44	82	4.6	803	1	SW16_YEAST	P09359 saccharomyc
45	81.5	4.6	300	1	GGPP_HUMAN	O95749 h geranylge

ALIGNMENTS

RESULT 1
ULP1_SCHPO
ID ULP1_SCHPO STANDARD; PRT; 568 AA.
AC Q42957:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiquitin-like-specific protease 1 (EC 3.4.22.-).
GN ULP1 OR SPBC1967.09.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkhardt G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesti D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
RN [2]
FUNCTION, AND GENE NAME.
RA Watts F.;
RA Submitted (MAR-2001) to the SWISS-PROT data bank.
CC -!- FUNCTION: A CYSTEINE PROTEASE THAT PROTEOLYTICALLY REMOVES THE
CC C-TERMINUS OF PMT3.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL; AL021839; CAAL17063.1; -
DR HSSP; Q02724; 1EUV.
DR MEROPS; C48.002; -
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
DR PROSITE; PSS0600; ULP_PROTEASE; 1.
KW Hydrolase; thiol protease.
FT ACT_SITE 463 463 BY SIMILARITY.
FT ACT_SITE 480 480 BY SIMILARITY.
FT ACT_SITE 527 527 BY SIMILARITY.
SQ SEQUENCE 568 AA; 64940 MW; AAF9771440302D9F CRC64;
Query Match 18.0%; Score 319.5; DB 1; Length 568;
Best Local Similarity 28.7%; Pred. No. 2.2e-18;
Matches 90; Conservative 51; Mismatches 114; Indels 59; Gaps 9;
QY 59 EKAGHSPLREHVTVCQS-----ILDEPLQYGSILPLSTDEVVEKLEDFQQ----- 107
Db 264 KSPHDSRKRASRFVQSDQPARNIIVTSDIQEKSLLLIIRD-LKEQTESFQDMNEV 322
QY 108 -----EFSTPSRKLGLVLIQTSQRMGPNAMVRGFRVAYKRH 144
Db 323 DFLQLGLEISPPTRPKFPELEFPDPAKRAKLYLNQSNVSSSEPIITKFNIP----- 378
QY 145 VLTWDDLTGLQWLNDOVMYGDIV-----MDTPVKVHFNSFFYKLRKTKGVDGV 199
Db 379 -ITLKDHLTNRLQWLNDEINFYNNLISERSKIDSLPRVHGFTFTYSLQRRGYAGV 437
QY 200 KRWTK--NVDIFNKELLIPHLHVBHWSLISVDVRRRTITVFDQSORTLNRCPKHIAYKL 257
Db 438 RRAKARVNIADDAVFIPLVHVMCAVINSKRFEWDSLAGSPGVFLDLRDY 497
QY 258 QAEAVKDRDLDFHOGWKYFKMVARQNDSDCGAFVLQYCKHALSQPFSFTQDDMKPL 317
Db 498 IAET--KGAVDV--SDWENFMDDNSPRQRNGHDCGVFACKTAECVSRNVPVQFSQNDMP 554
QY 318 R-----RQIY 322
Db 555 RIKMAASIIDAQIY 568
RESULT 2
YRW3_CAEL
ID YRW3_CAEL STANDARD; PRT; 659 AA.
AC Q09353;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein T10F2.3 in chromosome III.
GN T10F2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=Bristol N2.
RA Miller N.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST LPB11C AND S.POMBE SPBC19G7.09.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL; U23412; AAK21468.1; -
DR HSSP; Q02724; 1EUV.
DR MEROPS; C48.002; -
DR WormPep; T10F2.3; CE26951.
DR InterPro; IPR003653; SUMO_protease.
DR Pfam; PF02902; Peptidase_C48; 1.
DR PROSITE; PSS0600; ULP_PROTEASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 659 AA; 75304 MW; 4F66F0082D470EB7 CRC64;
Query Match 15.2%; Score 268.5; DB 1; Length 659;
Best Local Similarity 26.4%; Pred. No. 3.6e-14;
Matches 75; Conservative 53; Mismatches 79; Indels 77; Gaps 10;
QY 27 LISNVCISG---DHVAQELFQSGDLGMAEAPRGQKAGQHSPLREHVTVCQSILDFL 83
Db 393 ILSQINSLGIGSAYRGPRYQNS-----YQLSKQKEDKLLLEARIREGHS----- 438
QY 84 QTVGSLPLSTDEVVEKLEDFQOEFTPSRKLGLVLIQTSYQRM----- 128
Db 439 QTRG-----DRLEDV-----RKRLQGLAIAPKPKVKKKVDDEMALPDAA 478
QY 129 -----PCNAMVRGFRVAYKRHLVTMDLGLTYGNLNDQVMNMYGDLVM--- 173
Db 479 DALVERAMSGGNNEQVDAFSIQ-----ICKKDLATLSGLHNLNDEIIFYLQICDRS 533
QY 174 --DTVPEKVFHFFNSFFYDKLRTKGYCGVKKRWTKNVDIFNKELLIPHLHVBHWSLISVDV 231
Db 534 NGDSKYPKIYAFNTFFYSNIVSGYASVKKRWTKVDIFAFLVPLVHGLMHWCMVIDM 593
QY 232 RRTTITVFDQSORTLNRCPKHIAYLQAEAV--KKDLRDFHGW 273
Db 594 GKRIEYFSDYLDGNTAVLPALRGYLEAESLDDKKKTAMNF-SGW 636
RESULT 3
ULP1_YEAST
ID ULP1_YEAST STANDARD; PRT; 621 AA.
AC Q02724;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiquitin-like-specific protease 1 (EC 3.4.22.-).
GN ULP1 OR YPL020C OR LPB11C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sidu A.M., Tettelin H.,
RA Urrestazu L.A., Ushinsky S., Vierdeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";


```
FT DOMAIN 694 984
FT ACT_SITE 794 794
FT ACT_SITE 926 926
FT CONFLICT 275 275
FT CONFLICT 330 330
FT CONFLICT 546 546
FT CONFLICT 747 761
FT CONFLICT 828 828
FT CONFLICT 881 881
SQ SEQUENCE 984 AA: 112264 MW: 317E8B7B4128EDID CRC64;

Query Match 9.6%; Score 170.5; DB 1; Length 984;
Best Local Similarity 22.6%; Pred. No. 5.1e-06;
Matches 64; Conservative 41; Mismatches 67; Indels 111; Gaps 11;

QY 146 LTMDDLTLYGQWLNQVMNMYGDLV-----DIVPEKVHFFNFFDKLRTK----- 194
DB 696 VTNEDELECEGEFLNDVIIDFYLYKYLEKASDELVERSIFSSFFYKCLTRKENNLTE 755
QY 195 -----GYDGVKRWTKNDIENKELLPIPIHLEVHWSL----- 226
DB 756 DNPNLMSAQRRHKRVTRRHINENKDYIFVPVNESSHWYLAICFPWLEAVYEDFPQ 815
QY 227 -----ISVDVRRR---TITVFDQSORT-LNRRCPKHIAK----- 255
DB 816 TVSQSQSAQSQNDKNTIDNLTSTLSAEDSOSTESNMSPVKMKRCPICILIDSL 875
QY 256 -----YLQAEAVKDRIDFHQGWKGFKMN-----VARQNNDSCGAFV 294
DB 876 KAASQVNTVQNLRELYVEWEVK--LKTH---RQFSKTNMVDLCPKPKQDNSSDCGVYL 930
QY 295 LOYCKHLALSOPSTQODMP-----KLRRQIYKEL 325
DB 931 LQYVESFKDPVNF---ELPIHLEKWPFRHVIKTKREDIREL 970

RESULT 7
SUSL_HUMAN
ID SUSL_HUMAN STANDARD; PRT; 1112 AA.
AC Q9GZRL: Q9UJY5; O94891;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE SUMO-1-specific protease 1 (EC 3.4.22.-) (Sentrin-specific protease
DE SENP6) (Protease FKSG6).
GN SUSP1 OR SSP1 OR KIAA0797.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=20261527; PubMed=10799485;
RA Kim K.I., Baek S.H., Jeon Y.-J., Nishimori S., Suzuki T., Uchida S.,
RA Shimbara N., Saitoh H., Tanaka K., Chung C.H.;
RT "A new SUMO-1-specific protease, SUSP1, that is highly expressed in
RT reproductive organs."
RL J. Biol. Chem. 275:14102-14106(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20267842; PubMed=10806345;
RA Yeh E.T., Gong L., Kamitani T.;
RT "Ubiquitin-like proteins: new wines in new bottles."
RL Gene 248:1-14(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Wang Y.-G.;
RT "Identification of FKSG6, a novel protein with protease activity."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
```

```
RP SEQUENCE OF 29-1112 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 5:277-286(1998).
RC -!- FUNCTION: Protease that release SUMO-1 from its precursor
CC sequence.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Highly expressed in reproductive organs, such
CC as testis, ovary, and prostate.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF196304; AAF04852.1; -
CC EMBL; AF307849; AAG29831.1; -
CC EMBL; AF306508; AAG30253.1; -
CC EMBL; AB018340; BAA34517.1; -
CC MEROPS: C48.004; -
CC MIM: 605003; -
CC InterPro: IPR003653; SUMO-protease.
CC Pfam: PF02902; Peptidase_C48; 1.
CC PROSITE: PS0600; ULP_PROTEASE; 1.
KW Hydrolase; Thiol protease; Ubl conjugation pathway.
FT DOMAIN 666 1112
FT ACT_SITE 765 765
FT ACT_SITE 917 917
FT ACT_SITE 1030 1030
FT CONFLICT 121 121
FT CONFLICT 1043 1043
FT CONFLICT 1106 1106
SQ SEQUENCE 1112 AA; 126144 MW; B6E53875C3833A46 CRC64;

Query Match 8.1%; Score 143; DB 1; Length 1112;
Best Local Similarity 30.8%; Pred. No. 0.001;
Matches 32; Conservative 23; Mismatches 29; Indels 20; Gaps 3;

QY 146 LTMDDLTLYGQWLNQVMNMY-GDLVMDTV-----PEKVHFFNFFDKL----- 191
DB 668 VTNEDELECEGEFLNDVIIDFYLYKYLEKEDADRHIFFSFFYKRLNQRRRNHE 727
QY 192 -----RTKGYDGVKRWTKNDIENKELLPIPIHLEVHWSLISV 229
DB 728 TTNLSIQQRHGRVKTWTRHVDIFEKDFIFVPLNEAAHFWLAV 771

RESULT 8
YOG6_CAEEL
ID YOG6_CAEEL STANDARD; PRT; 342 AA.
AC Q09275;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 35.7 kDa protein C41C4.6 in chromosome II.
GN C41C4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Burton J.;
```



```
QY 75 VOSILDEFLQYGS-----IPLSTDEVVEKLEDFQEFST- 111
  ||:|||||:|:|
Db 293 IQNALDEFLPHHSIDDKYFQEWETSDKSLDVPDCRIDLSVFNWQSSENCYEPFKTG 352
QY 112 --PSRKGVLVLQISQYQRMGPNVGRFVAYKRVHVLMDLGLTLYGQNLNDQVNMVYG 169
  ||:|||||:|:|
Db 353 ALSTRKGTQTEAL-----LAIKRNMVNPNGIYDYNVSVANSVVKLL 396
QY 170 DLVMTVPKGVHFFNSF-----FYDKLRKGVGVKRWTKNVD 207
  ||:|||||:|:|
Db 397 TTVID--FDKLCMFDFISEGVSFQDYIVGKNPDPELSDPLGVSIDSYKHKMSV- 453
QY 208 IFNKELLIP-----HLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIKAKYLQAEAV 262
  ||:|||||:|:|
Db 454 -----LKPVEDNSLHLE-----RMPATIIYHKDKIWS--SSPIFLAAARLMJI 497
QY 263 KKDRL-----DPHQ 271
  ||:|
Db 498 LRDKITIPSGKPHQ 511

RESULT 11
VPRT_ASFB7
ID VPRT_ASFB7 STANDARD; PRT; 273 AA.
AC Q00946;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Polyprotein processing peptidase (EC 3.4.22.-).
GN S273R.
OS African swine fever virus (strain BA71V) (ASFV).
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
OX NCBI_TaxID=10498;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92263807; PubMed=1316688;
RA Garcia-Beato R., Freije J.M., Lopez-Otin C., Blasco R., Vinnuela E.;
RT "A gene homologous to topoisomerase II in African swine fever virus.";
RL Virology 188:938-947(1992).
RN [2]
RP COMPLETE GENOME.
RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
RA Rodriguez J.F., Vinnuela E.;
RT "Analysis of the complete nucleotide sequence of African swine fever
  virus.";
RL Virology 208:249-278(1995).
RN [3]
RP FUNCTION, CHARACTERIZATION, AND MUTAGENESIS.
RX MEDLINE=20576431; PubMed=11031264;
RA Andres G., Alejo A., Simon-Mateo C., Salas M.L.;
RT "African swine fever virus protease, a new viral member of the
  SUMO-1-specific protease family.";
RL J. Biol. Chem. 276:780-787(2001).
CC -!- FUNCTION: Protease involved in the processing of the virus
  polyproteins.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88699; AAA42734.1; -.
CC EMBL; U18466; AAA65340.1; -.
CC PIR; A42549; A42549.
CC MEROPS; C48.050; -.
KW Hydrolase; Thiol protease; Late protein.
FT ACT_SITE 168 168 PROBABLE.
```

```
FT ACT_SITE 187 187 POTENTIAL.
FT ACT_SITE 232 232 PROBABLE.
FT MUTAGEN 168 168 H->R: LOSS OF ACTIVITY.
FT MUTAGEN 232 232 C->S: LOSS OF ACTIVITY.
SQ SEQUENCE 273 AA; 31550 MW; C5AA778179C57C17 CRC64;

Query Match 5.2%; Score 92.5; DB 1; Length 273;
Best Local Similarity 25.9%; Pred. No. 2.1;
Matches 29; Conservative 20; Mismatches 44; Indels 19; Gaps 5;

QY 223 HWSLISVDVRR--TITYFDSQRTLNRRCPKHIKAKYLQAEAVKRLDFHQGWKGYFKNM 280
  ||:|||||:|:|
Db 168 HWAIVFDMRGDCWSEYFNS---AGNSPPGPVIRWM--ERVQQLLKIHHTVKTAVTN 222
QY 281 VARQNNDSCGAFVLOYCK-----HLALSQPFSTQODMPKLRRRIYK 323
  ||:|||||:|:|
Db 223 INHQRSQTECGPSLYFIRARLDNVSYAHFISAR---ITDEMYKERTHLFR 271

RESULT 12
YR22_CAEEL
ID YR22_CAEEL STANDARD; PRT; 727 AA.
AC Q09639;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative G protein-coupled receptor kinase W02B3.2 (EC 2.7.1.-).
GN W02B3.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Connell M.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC GPRK SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 1 RGS DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U22833; AAA64322.1; -.
CC HSSP; P25098; IBAK.
CC WormPep; W02B3.2; CE02074.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR000239; GPCR_Kinase.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000961; Pkinase_C.
CC InterPro; IPR000342; Regl_Gproteins.
CC InterPro; IPR002290; Ser_thr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00615; RGS; 1.
CC PRINTS; PR00717; GPCR_KINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00233; PH; 1.
CC SMART; SM00315; RGS; 1.
CC SMART; SM00133; S_TK_X; 1.
CC SMART; SM00220; S_TK; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS00132; RGS; 1.
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
  ATP-binding.
FT DOMAIN 1 190 N-TERMINAL.
```

```
FT DOMAIN 191 455 PROTEIN KINASE.
FT DOMAIN 456 700 C-TERMINAL.
FT DOMAIN 54 175 RGS.
FT NP_BIND 197 205 ATP (BY SIMILARITY).
FT BINDING 220 220 ATP (BY SIMILARITY).
FT ACT_SITE 318 318 BY SIMILARITY.
SQ SEQUENCE 727 AA; 83361 MW; AA05308A9DA57D7 CRC64;

Query Match 5.1%; Score 90; DB 1; Length 727;
Best Local Similarity 19.7%; Pred. No. 11;
Matches 62; Conservative 50; Mismatches 122; Indels 80; Gaps 15;

QY 48 LGMAEAEPEKAGQSPHREHVTCVQSILDEFLQTYGSLPLSTDEVVEKLEDFQ 107
Db 15 MAMEKRSQPAARAKRVLDPD--PSVRSIMQKFELEKSGDM-----KFDKIFNQ 61
QY 108 EF-----STPSKGLVLQIQSYQRM--PCNAMVRGFRVAYKRHLVLTMDL 152
Db 62 KLGFLLLKDYAENVSESPQIKFYEAKEYKMETPDERLTKA-REIYDHHIM-VEMLA 119
QY 153 TLVGONLNDQVMNMGDLVMDTV-PEKVHFFNSFFYDKLRTGKYGDKRWTKNVDIFNK 211
Db 120 --HANVSKESLQVHLLKQNVPPDLFRHYVLEICDQLRG-----DIFQR 164
QY 212 ELLIPIHLEHVHWSLISVDVRRRTITYFDSQRTLN-----RCPK-HIAKYLQAEAVK 263
Db 165 FLESKDTRFCQWKNLELN--QLTMNDFSVHRIIGRGGEVYGCRAKDTGKMYAMKCLD 223
QY 264 KRLDFHQHKGVEK-----MNVARQNNDSDCGAFVLQYCK-----H 300
Db 224 KRKKMKGQETLALNEHIMLSLSTGDCPFIVCMTYAFQSPDKLC--FILDLMNGDGLH 281
QY 301 LALSQPFSTQDDM 314
Db 282 YHLSQHGVTQEOM 295

RESULT 13
HTPG_CAMJE STANDARD; PRT; 608 AA.
AC QSPH23;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chapterone protein htpg (Heat shock protein htpg) (High temperature protein G).
DE HTPG OR CJ0518.
GN Campylobacter jejuni.
OS Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karylishev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL139075; CAB75155.1; -.
DR HSSP; P02829; IAH8.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR001404; Hsp90.
DR Pfam; PF00183; HSP90; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR PRINTS; PR00775; HEATSHOCK90.
DR SMART; SM00387; HATPase_C; 1.
DR PROSITE; PS00298; HSP90; 1.
KW Chapterone; ATP-binding; Heat shock; Complete proteome.
FT DOMAIN 1 332 A; SUBSTRATE-BINDING (BY SIMILARITY).
FT DOMAIN 333 536 B (BY SIMILARITY).
FT DOMAIN 537 608 C.
SQ SEQUENCE 608 AA; 69654 MW; 7FB40F4B365DF6B5 CRC64;

Query Match 5.0%; Score 89; DB 1; Length 608;
Best Local Similarity 21.1%; Pred. No. 11;
Matches 62; Conservative 45; Mismatches 95; Indels 92; Gaps 14;

QY 68 REBHTVCQSIL--DEFLQTYGSLPLSTDEVVEKLEDFQ-----QEFSTPSRKG-- 116
Db 160 KERQGTSTILYKXDEFANAY-----KIESIIEKYSNHIQFFIFMEKEEF-TPAKEGEE 212
QY 117 ---LVQLQIQSYQRMPCNAMVRGFRVAYKRHLVLTMDLGLTYGONLNDQVMNMGDLVM 173
Db 213 EGTEKLIQSINK--ANALWR-----MOKSLKAEDYERFYEDNFHDSNKKPLLY---L 260
QY 174 DTVPKRVHFFNSFF-----YDKLRTGKYGDKRWTKNVDIFNKELLIPIHLEHVHWSL 226
Db 261 HTKSEKLEYNLSLFFIPQNAFDFLRVDYQSGLKLVKRVFISDDDKELLPTVLRFRVGI 320
QY 227 ISVDVRRRTITYFDSQRTLNRCPKHIAKYLQAEAVKORLDPHQHKGVEKNNVAR--- 283
Db 321 IDVE-----DLPLN-----VSREILOENQILKGIKEASVKKILG 354
QY 284 -----QNNDSDCGAFVLQYCKHLALSQPF-----SFTQDDMPKLRRIQYK 323
Db 355 ELEKLNNDKE-----KYLSPFTFGKVLKGLYGGGKDSLLKLMLYK 399

RESULT 14
VPRT_ASEF2 STANDARD; PRT; 273 AA.
AC Q65228;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Polypeptide sequence of a 55 kbp region from the right end of the
DE ORF16R.
GN ORF16R.
OS African swine fever virus (isolate Malawi Lil 20/1) (ASFV).
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
OX NCBI_TaxID=10500;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94292916; PubMed=8021596;
RA Dixon L.K., Twigg S.R.F., Baylis S.A., Vydelingum S., Bristow C.,
RA Hammond J.M., Smith G.L.;
RT "Nucleotide sequence of a 55 kbp region from the right end of the
RT genome of a pathogenic African swine fever virus isolate (Malawi
RT LIL20/1)."
RL J. Gen. Virol. 7:1655-1684(1994).
CC -1- FUNCTION: Protease involved in the processing of the virus
CC polyproteins (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL; X71982; CAA50819.1; -;
DR MEROPS; C48.050; -;
KW Hydrolase; Thiol protease; Late protein.
FT ACT_SITE 168 187 BY SIMILARITY.
FT ACT_SITE 187 187 BY SIMILARITY.
FT ACT_SITE 232 232 BY SIMILARITY.
SQ SEQUENCE 273 AA; 31493 MW; D6AD6DAC62C25F3B CRC64;

Query Match 4.9%; Score 87.5; DB 1; Length 273;
Best Local Similarity 25.0%; Pred. No. 5.2;
Matches 28; Conservative 21; Mismatches 44; Indels 19; Gaps 5;

QY 223 HNSLISVDVRR--TTTFDSQRTLNRCPKHIAYQLQAEVKKRDLDFHQGKGFKNW 280

Db 168 HWAIFVDMRGDCWSIEYFNS---AGNPPGPVIRWM--ERVQQLKIHHTVKTIAVTN 222

QY 281 VARQNDSDCGAFVLYCK-----HLALSQPSFTQDMPLRROIYK 323

Db 223 IRQRSQTECGPYSLFYIRALDNVSYTHFISTR---ITDENMKPRTHLFR 271

RESULT 15

DNLJ_MYCLE
ID DNLJ_MYCLE STANDARD; PRT; 694 AA.
AC O33102;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA ligase (EC 6.5.1.2) (Polydeoxyribonucleotide synthase [NAD+]).
GN LIGA OR Lig OR ML1705 OR MLCB637.10.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Muncall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus";
RL Nature 409:1007-1011(2001).

CC -!- FUNCTION: THIS PROTEIN CATALYZES THE FORMATION OF PHOSPHODIESTER
CC LINKAGES BETWEEN 5'-PHOSPHORYL AND 3'-HYDROXYL GROUPS IN DOUBLE-
CC STRANDED DNA USING NAD AS A COENZYME AND AS THE ENERGY SOURCE FOR
CC THE REACTION. IT IS ESSENTIAL FOR DNA REPLICATION AND REPAIR OF
CC DAMAGED DNA (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: NAD(+) + [deoxyribonucleotide](N) +
CC [deoxyribonucleotide](M) = AMP + nicotinamide nucleotide +
CC [deoxyribonucleotide](N+M).
CC -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT DNA LIGASE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 BRCT DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL; Z99263; CAB16425.1; -;
DR EMBL; AL583923; CAC30658.1; -;
DR HSP; O87703; 1B04.
DR Leproma; ML1705; -;
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR004150; DNA_ligase_OB.
DR InterPro; IPR001679; DNALigase.
DR InterPro; IPR003583; HHH_1.
DR InterPro; IPR000445; Hhh.
DR InterPro; IPR004149; Znf_DNAligase_C4.
DR Pfam; PF00533; BRCT; 1.
DR Pfam; PF00633; HHH; 1.
DR Pfam; PF01653; DNA_ligase_N; 1.
DR Pfam; PF03119; DNA_ligase_ZBD; 1.
DR Pfam; PF03120; DNA_ligase_OB; 1.
DR ProDom; PD003944; DNAligase; 1.
DR SMART; SM00292; BRCT; 1.
DR SMART; SM00278; HHH1; 1.
DR SMART; SM00532; LIGANC; 1.
DR TIGRFAMS; TIGR00575; dnlj; 1.
DR PROSITE; PS0172; BRCT; 1.
DR PROSITE; PS01055; DNA_LIGASE_N1; 1.
DR PROSITE; PS01056; DNA_LIGASE_N2; FALSE_NEG.
KW Ligase; DNA repair; DNA replication; NAD; Complete proteome.
FT DOMAIN 613 677 BRCT.
FT BINDING 123 123 AMP (BY SIMILARITY).
SQ SEQUENCE 694 AA; 75912 MW; CFE712B6AFDAA9D5 CRC64;

Query Match 4.9%; Score 87.5; DB 1; Length 694;
Best Local Similarity 23.1%; Pred. No. 17;
Matches 67; Conservative 37; Mismatches 123; Indels 63; Gaps 13;

QY 45 GSDLGNA-----EAEERPGKAGCOHSPLEEHVTCVQSILDELOTYQSLPLSTDEVVE 99

Db 360 GSTVGQATLHNPSEVRKRGVLGDTVVIRKA-----GDVPEVLGPVVD 403

QY 100 KLEDIPQOEF-----STPSRKGLVLQLIQSYORMPGNAMVGRFVAYKRHLVT 147

Db 404 -LRDGSEREFEVMTTCPEGCTTLAPEKEGDADIRCFNARSCPQLRERVFHVA-SRSALD 461

QY 148 MDDLGLTYGQNLNDQVMNMGDLVMDTVPEKVFHFFNSFFYDKLRTKGYDGVKRWTKNVD 207

Db 462 IOGLGYEAGVALLAAQVITSEGLF--TLTEKALLRTELFRNKAGELSANG-KRLLVNVD 518

QY 208 IFNKELLPIHLEVHWS-LISVDVR-----RRITVFDQSORTLNRRCPKHIAY--- 256

Db 519 -----KAKTAPL-----WRVLVALSIRHVGPRTAARALATEFGSDAILAASPEQLAAEGV 569

QY 257 ---LQAEAYKKDRLDPHQGWKGYFRKMNVARQNNDSDCGAFVLYQYKHLAL 303

Db 570 GTTIAAAVTEWTFVDWHRVIVNKNRAAGVGMVDERDTS--VLPTCBGLTI 617

Search completed: December 16, 2002, 19:58:11
Job time : 210 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2002, 20:15:50 : Search time 299 seconds
(without alignments)
2500.547 Million cell updates/sec

Title: US-09-848-852A-3
Perfect score: 1772
Sequence: 1 MYSQREWTGWIARRGAHLA.....DMPKLRQIKYELCHCKLTV 332

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q/cgn2.1/USPFO-spool/US09848852/runat_16122002_132039_19101/app-query.fasta_1.519
-DB=N.Geneseq_101002 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09848852 -CGCN_1.1.281 @runat_16122002_132039_19101 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq_101002.*
1: /SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1751	98.8	1760	21	AAZ43782 Human fetal brain
2	1751	98.8	2080	22	AAI58425 Human polynucleoti
3	1751	98.8	2206	22	AAF54831 Nucleotide sequenc
4	1703	96.1	1507	22	AAZ05541 Human secreted pro
5	1702	96.0	1934	22	AAZ05502 Human secreted pro
6	1699	95.9	1804	22	AAI60211 Human polynucleoti
7	1699	95.9	1804	22	AAI60212 Human polynucleoti
8	1674	94.5	1933	24	ABQ54216 Human ovarian anti
9	1642	92.7	2029	22	AAI58426 Human polynucleoti
10	820	46.3	1578	22	AAH13708 Human cDNA sequenc
11	784	44.2	870	22	AAI94288 Human neuroblastom
12	755.5	42.6	1938	22	AAH99558 Human protein enco
13	661.5	37.3	503	22	AAZ52529 Human ovarian PCR-
14	618.5	34.9	358	22	AAI84371 Human polynucleoti
15	614	34.7	820	22	AAI93958 Human neuroblastom
16	580	32.7	714	22	AAH03368 Human cDNA clone (
17	474.5	26.8	5318	22	AAF55961 Human eukaryotic 1
18	456	25.7	2512	22	AAF54827 Nucleotide sequenc
19	393	22.2	2570	23	ABV22135 Human prostate exp
20	393	22.2	2570	23	ABV27975 Human prostate exp
21	393	22.2	2795	22	AAF54830 Nucleotide sequenc
22	393	22.2	3133	22	AAI93888 Human stomach canc
23	393	22.2	3133	22	AAH17779 Human cDNA sequenc
24	353	19.9	650	21	AAC59569 Human secreted pro
25	312	17.6	1773	23	ABL05231 Drosophila melanog
26	311	17.6	3963	23	ABL05230 Drosophila melanog
27	310.5	17.5	9096	23	ABL17142 Drosophila melanog
28	308	17.4	7494	23	ABL13550 Drosophila melanog
29	307.5	17.4	595	20	AAV84693 Human EST Showing
30	307	17.3	4907	23	ABL13551 Drosophila melanog
31	299	16.9	181	21	AAC08625 Human secreted pro
32	297	16.8	666	23	AAZ71205 DNA encoding novel
33	289	16.3	208	21	AAZ45202 Human secreted exp
34	261.5	14.8	1859	21	AAC48444 Arabidopsis thalia
35	259.5	14.6	1861	21	AAC40619 Arabidopsis thalia
36	255.5	14.4	862	21	AAA02686 Human colon cancer
37	236	13.3	10713	23	ABL21052 Drosophila melanog
38	236	13.3	10734	23	ABL21050 Drosophila melanog
39	232	13.1	13525	23	ABL04986 Drosophila melanog
40	215.5	12.2	695	22	AAZ05540 Human secreted pro
41	205	11.6	631	20	AAV84692 Rice EST showing h
42	196	11.1	33954	21	AAZ93712 F-box protein Met3
43	190	10.7	501	22	AAK61344 Human immune/haem
44	173.5	9.8	3447	23	ABL16117 Drosophila melanog
45	170	9.6	2618	24	ABL41323 Human protein phos

ALIGNMENTS

RESULT 1

AAZ43782

ID AAZ43782 standard; cDNA; 1760 BP.

XX AAZ43782;

AC AAZ43782;

XX 10-MAR-2000 (first entry)

DT Human fetal brain cDNA clone vb7_1.

DE Human; secreted protein; treatment; nutritional activity; cytokine;
XX cell proliferation; cell differentiation; hematopoiesis regulation;
KW tissue growth; activin; inhibin; chemotactic; chemokinetic; hemostatic;
KW thrombolytic; anti-inflammatory; invasion suppressor; tumor inhibition;
KW gene therapy; ds.

OS Homo sapiens.

XX

PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; AAM39269.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1: SEQ ID NO 628; 10078pp; English.
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activ/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2080 BP; 472 A; 567 C; 570 G; 471 T; 0 other;

Alignment Scores:
 Pred. No.: 1,51e-193 Length: 2080
 Score: 1751.00 Matches: 331
 Percent Similarity: 99.40% Conservative: 0
 Best Local Similarity: 99.40% Mismatches: 1
 Query Match: 98.81% Indels: 1
 DB: 22 Gaps: 0

US-09-848-852A-3 (1-332) x AAI58425 (1-2080)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20
 DB 634 ATGTACTCTGCCACGGTTTGGGGCAATCTGGCCAGAGGGGCGCGCTGGC 693
 QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
 DB 694 ACCCCCTGATGCCAGCATCTCATCAGCAATGTGTGCAGCATCGGGACCCTGTGGCCCA 753
 QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60
 DB 754 GGAGCTTTTTCAGGGCTCAGATTGGGCATGGCAGAAAGGAGGAGCGCTGGGGAGAA 813
 QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
 DB 814 ACCCGCCAGACACGCCCTCGGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTGA 873
 QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValAlaGluLy 100
 DB 874 CGAATTCTTCAACGATATGGCAGCCTCATACCCCTCAGCACTGATGAGGTAGAGAA 933

QY 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGl 120
 DB 934 GCTGGAGGACATTTCCAGCAGAGAGTTTCCACCCCTCCAGAGAGGGGCTGGTGTGA 993
 QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
 DB 994 GCTGATCCAGTCTTACCAGCGGATGCCAGCAATGCCATGGTGGAGGGGCTCCGAGTGC 1053
 QY 140 aTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160
 DB 1054 TTATAAGCGGCACGTGCTGACCATGGATGGGACCTTGTATGGACAGCACTGGCT 1113
 QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
 DB 1114 CAATGACCCAGGTGATGACATGTATGGAGACCTGGTCATGGACACAGTCCCTGAAAAGT 1173
 QY 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLy 200
 DB 1174 GCATTTCTTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGTGA 1233
 QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuLeuIleProIleHisLe 220
 DB 1234 AAGGTGGACCAAAACGCTGGACATCTCAATAAGGAGCTACTGCTAAATCCCATCCACCT 1293
 QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgArgArgThrIleThrTyrPheAs 240
 DB 1294 GGAGGTGCATTTGGTCCCTCATCTCTGTGATGTAGGCGACGACCATCATTTTGA 1353
 QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260
 DB 1354 CTCGACGCTACCTTAAACCGCGCTGCCTAAGCATATTTGCCAAGTATCTACAGGCAGA 1413
 QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280
 DB 1414 GCGGTAAAGAAGACCGCATGCTGATTTCCACGAGGCTGGAAGGTTACTTCAAAATGA 1473
 QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHi 300
 DB 1474 TGTGGCCAGGCAGATAATGACAGTGCATGTGTGCTTTTGTGTGTCAGTACTGCAAGCA 1533
 QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgGl 320
 DB 1534 TCTGGCCCTGTCTCAGCCATTCAGCTTCCACGAGGAGCATGCCCAAACCTTCGTGCGCA 1593
 QY 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
 DB 1594 GATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTG 1630
 RESULT 3
 AAF54831
 ID AAF54831 standard; DNA; 2206 BP.
 XX
 AC AAF54831;
 DT 15-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of a human SENP2 polypeptide.
 XX
 KW Human; de-sentrinase; SENP; SENP1; SENP2; SENP3; chromosome 12q13.1;
 KW sentrin-specific protease; sentrin; anti-proliferative agent;
 KW anti-viral agent; herpes simplex-1; cytomegalovirus; sentrinisation;
 KW PML; tumour suppressor; acute promyelocytic leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 52..1758
 FT /*tag= a
 FT /product= *SENP3*
 XX
 PN WO200109292-A2.
 XX

PD 08-FEB-2001.
 XX
 XX 31-JUL-2000; 2000WO-US20884.
 XX
 XX 31-JUL-1999; 99US-0146774.
 XX
 XX (TEXA) UNIV TEXAS SYSTEM.
 XX
 XX Gong L, Yeh ETH;
 XX
 XX WPI; 2001-182947/18.
 DR P-PSDB; AAB31977.
 XX
 XX New polynucleotide encoding de-sentrinase polypeptides, useful as
 PT antiproliferative or antiviral agents -
 XX
 XX Claim 36; Page 120; 122pp; English.
 XX
 CC The present sequence encodes a human de-sentrinase (SENP) 3 polypeptide.
 CC The specification describes SENP1 and SENP2 polypeptides. The SENP1
 CC gene is found in chromosome 12q13.1. SENP1 is found in the nucleus,
 CC SENP3 is found in the nucleolus, and SENP3 is found in both locations.
 CC SENP polypeptides are sentron-specific proteases that remove sentrin
 CC from some sentrinised peptides, but do not affect proteins modified by
 CC ubiquitin or NEDD8. SENP polypeptides are used to identify specific
 CC modulators of SENP. These modulators are potential anti-proliferative
 CC and anti-viral agents such as herpes simplex-1 or cytomegalovirus.
 CC SENP polypeptides are used for studying the role of sentrinisation in
 CC the biological function of PML, a tumour suppressor implicated in
 CC development of acute promyelocytic leukaemia. Fragments of SENP
 CC polynucleotides are used as hybridisation probes and amplification
 CC primers for detecting gene expression or preparing mutated sequences.
 CC also as antisense sequences for inhibiting SENP expression.
 XX
 SQ Sequence 2206 BP; 491 A; 612 C; 605 G; 498 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.64e-193 Length: 2206
 Score: 1751.00 Matches: 331
 Percent Similarity: 99.40% Conservative: 0
 Best Local Similarity: 99.40% Mismatches: 1
 Query Match: 98.81% Indels: 1
 DB: 22 Gaps: 0
 US-09-848-852A-3 (1-332) x AAF54831 (1-2206)
 QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHis-LeuAl 20
 DB 759 ATGTACTCTGCCCAACGGTTTGGGGGACAAATCTGGGCCAGAGGGGAGCGCAGCTGGC 818
 QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1 40
 DB 819 ACCCCCTGATGCCAGCATCTCATCAGCATGTGTGCAGCATCGGGGACCATGTGGCCCA 878
 QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60
 DB 879 GGAGCTTTTTCAGGGCTTCAGATTGGCGATGGCAGAGAGGAGGAGGCGCTGGGGAGAA 938
 QY 60 sAlaGlyGlnHisSerProLeuArgGluHisValThrCysValGlnSerIleLeuAs 80
 DB 939 AGCGGGCCAGCAGACCCCTCGGAGAGGAGCATGTGACCTGGCTGACAGAGCATCTTGA 998
 QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100
 DB 999 CGAATTCCTTCAACGTATGGAGCCTCATACCCCTCAGCAGCTGATGAGTAGAGAA 1058
 QY 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuG1 120
 DB 1059 GCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGGCCTGGTTGCA 1118
 QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
 DB 1119 GCTGATCCAGCTTACCAGCGGATGCCAGCAATGCCATGTGGTGGGGGCTTCCGAGTGGC 1178

QY 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160
 DB 1179 TTATAAGCGCAGCTGCTGACCATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCT 1238
 QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
 DB 1239 CAATGACCAAGTGTATGAACATGTATGGAGACCTGGTGCATGGACACAGTCCCTGAAAGGT 1298
 QY 180 lHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLy 200
 DB 1299 GCATTTCTTCAATAGTTTCTTCTATGATAAATCCGTACCAAGGGTATGATGGGTGAA 1358
 QY 200 sArgTyrThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220
 DB 1359 AAGTGGACCAAAAACGTGGACATCTTCAATAAGGAGCTACTGCTAATCCCCATCCACCT 1418
 QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
 DB 1419 GGAGGTGCATTGCTCCCTCATCTCTGTGTATGTGAGCGACGACCATCATCTATTTTGA 1478
 QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaG1 260
 DB 1479 CTCGACGCTTACCTAAACGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAGA 1538
 QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280
 DB 1539 GCGGGTAAAGAAAGACCGACTGGATTTCACACAGGCGTGGAAAGGTACTTCAAAATGAA 1598
 QY 280 nValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysH1 300
 DB 1599 TGTGGCCAGGCAGCAATAATCACAGTACTGTGTGCTTTTGTGTCAGTACTGCAAGCA 1658
 QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgG1 320
 DB 1659 TCTGGCCCTGTCTCAGCCATTTCAGCTTACCCAGCAGGACATGCCAAACTTCGTCGGCA 1718
 QY 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
 DB 1719 GATCTACAGGAGCTGTGTCTACCTGCAAACTCAGCTGTG 1755
 RESULT 4
 AAD05541/c
 ID AAD05541 standard; cDNA; 1507 BP.
 XX
 AC AAD05541;
 XX
 DT 18-JUL-2001 (first entry)
 XX
 DE Human secreted protein-encoding gene 11 cDNA clone HOF0B27, SEQ ID NO:60.
 DE
 KW Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
 KW endocrine disorder; infection; wound healing; vulnerability;
 KW cell culture; chemotaxis; food additive;
 KW binding partner identification; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 CDS 416..715
 FT /*tag= a
 FT /product= "Human secreted protein precursor"
 FT /note= "Does not include start codon"
 FT /partial
 FT sig_peptide 416..493
 FT /*tag= b

FT mat_peptide 494..712
 FT /*tag= c
 FT /product= "Mature human secreted protein"
 PN WO200134767-A2.
 XX 17-MAY-2001.
 PD 01-NOV-2000; 2000WO-US30036.
 XX 05-NOV-1999; 99US-0163576.
 PR 27-JUL-2000; 2000US-0221366.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Soppet DR, Komatsoulis G, Shi Y, Olsen HS, Ruben SM;
 XX WPI; 2001-316492/33.
 DR P-PSDB; AAE01721.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition
 XX
 PS Claim 1; Page 447; 540pp; English.
 XX
 CC AAD05492-AAD05564 represent cDNAs corresponding to 22 human secreted
 CC protein genes, and AAE01672-AAE01743 represent the proteins they encode.
 CC AAE01744-AAE01763 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 22 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein-encoding cDNA of the invention.
 XX
 SQ Sequence 1507 BP; 357 A; 385 C; 354 G; 401 T; 10 other;
 SQ
 Alignment Scores:
 Pred. No.: 3,68e-188 Length: 1507
 Score: 1703.00 Matches: 320
 Percent Similarity: 99.69% Conservative: 1
 Best Local Similarity: 99.38% Mismatches: 1
 Query Match: 96.11% Indels: 0
 DB: 22 Gaps: 0
 US-09-848-852A-3 (1-332) x AAD05541 (1-1507)
 QY 11 IleTPrAlaArgGlyAlaHisLeuAlaProAspAlaSerIleLeuIleSerAsn 30
 DB 1453 CTTTGGCCAGAGGGGAGCGGANTTGGCACCCTGATGCCAGCATCTCTATCAGCAAT 1394
 QY 31 ValCysSerIleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAspLeuGlyMet 50
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 1393 GTGTGCAGCATCGGGGACCATGTGCCAGGAGCTTTTTCAGGGCTCAGATTTGGGCATG 1334
 QY 51 AlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHisSerProLeuArgGluGlu 70
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1333 GCAGAAGAGGCAGAGAGGCTGGGAGAAAGCCGCCAGCACAGCCCCCTGGCAGAGGAG 1274
 QY 71 HisValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySerLeuIle 90
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1273 CATGTGACCTGGTACAGAGCATCTTGGACGAATTCCTTCAACAGTATGGCAGCCTCAT 1214
 QY 91 ProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGlnIlePheSer 110
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1213 CCCCTCAGCACTGAYGAGGTAGTAGAARATRGAGGACATTTTCAGCAGAGATTTCY 1154
 QY 111 ThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMetProGly 130
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1153 ACMCTTCAGGAAGGGCTGGTGTGAGCTGATCCAGTCTACAGCGGATGCCAGC 1094
 QY 131 AsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThrMetAspAsp 150
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1093 AATGCCATGGTGAGGGCTTCCGAGTGGCTTATAAGCGGCACGTGCTGACCATGGATGAC 1034
 QY 151 LeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrGlyAsp 170
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1033 TTGGGGACCTTGTATGGACAGAACTGGCTCAATGACCCAGGTGATGAACATGATGGAGAC 974
 QY 171 LeuValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyrAspLys 190
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 973 CTGGTCATGGACACAGTCCCTGAAAGGTGCATTTCTCAATAGTTTCTTCATGATAAA 914
 QY 191 LeuArgThrLysGlyTyrAspGlyValLysArgTyrThrLysAsnValAspIlePheAsn 210
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 913 CTCCTGATCCAGGGTTATGATGGGGTGAAGAGTGGACCAAAAACGTGGACATCTCAAT 854
 QY 211 LysGluLeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIleSerValAsp 230
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 853 AAGGAGCTACTGCTAATCCCATCCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTGAT 794
 QY 231 ValArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgArgCysPro 250
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 793 GTGAGCGCAGCGACCATCACCTATTTTGACTCGCAGCGTACCCTAAACCGCGCTGCCCT 734
 QY 251 LysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHis 270
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 733 AAGCATATTGCCAAGTACTACAGCAGAGCGGTAAAGAAAGACCGAGTGGATTTCCAC 674
 QY 271 GlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAsnAsnAspSerAspCys 290
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 673 CAGGGCTGGAAAGGTACTTCAAAATGAATGTGGCCAGGCAGAGATAATGACAGTACTGT 614
 QY 291 GlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerGlnPropheSerPheThr 310
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 613 GGTGCTTTTGTGTCAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCAGCTTCACC 554
 QY 311 GlnGlnAspMetProLysLeuArgArgGlnIleTyrLysGluLeuCysHisCysLysLeu 330
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 553 CAGCAGGACATGCCCAAACTTCGTCGCCAGATCTACAGGAGCTGTGTCACATGCAAACTC 494
 QY 331 ThrVal 332
 |||||||
 Db 493 ACTGTG 488
 RESULT 5
 AAD05502
 ID AAD05502 standard; cDNA; 1934 BP.
 XX
 AC AAD05502;
 XX
 DT 18-JUL-2001 (first entry)
 XX
 DE Human secreted protein-encoding gene 11 cDNA clone HOF0B27, SEQ ID NO:21.
 XX
 KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;

KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;
 KW angogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KW cell culture; chemotaxis; vulvulery; binding partner identification;
 KW gene therapy; ss.

OS Homo sapiens.

Key	Location/Qualifiers
CDS	441..713
FT	/*tag= a
FT	/product= "Human secreted protein precursor"
FT	441..521
FT	/*tag= b
FT	522..710
FT	mat_peptide
FT	/*tag= c
FT	/product= "Mature human secreted protein"

PN WO200134767-A2.

17-MAY-2001

01-NOV-2000. 2000WQ-HIS2003E
XX
BEXX
DB
0E-NOV-1000-
0000-0103576

PR 27-JUL-2000; 2000US-0221366.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Soppet DR, Komatsoulis G, Shi Y, Olsen HS, Ruben SM;

DR WPI; 2001-316492/33.

XX Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -

PS Claim 1; Page 417-418; 540pp; English.

AAO05492-AAO05564 represent cDNAs corresponding to 22 human secreted protein genes, and AAE01672-AAE01743 represent the proteins they encode. AAE01744-AAE01763 represent human secreted protein fragments or variants. The secreted proteins and their genes are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 22 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, celiac disease, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infectious diseases. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

Db 1396 CCTGGCCCTGTCTCAGCCATTGAGTTCACCCAGGAGCATGCCCAAACTCGCCGGCA 1455
|||||
QY 320 nileTyrLysGluLeuCysHisCysLysLeuThrVal 332
|||||
Db 1456 GATCTACAAGGAGCTGTGTCACTGCAAACTCACTGTG 1492
|||||
RESULT 6
AAI60211/c
ID AAI60211 standard; cDNA; 1804 BP.
XX AAI60211;
AC AAI60211;
XX AAI60211;
DT 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4200.
DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dermanac RT;
XX
DR WPI; 2001-442253/47.
XX P-PSDB; AAM41055.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 4200; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activity, chemotactic/chemokinetic activity, haemostatic
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
XX specification.
SQ Sequence 1804 BP; 431 A; 477 C; 468 G; 428 T; 0 other;

Alignment Scores: 1.4e-187 Length: 1804
Pred. No.: 1699.00 Matches: 331
Score: 98.22% Conservative: 0
Percent Similarity: 98.22% Mismatches: 1
Best Local Similarity: 95.88% Indels: 5
Query Match: 22 Gaps: 0
DB: 22
US-09-848-852A-3 (1-332) x AAI60211 (1-1804)
QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20
|||||
Db 1485 ATGTACTCTGCCCAACGGTTTTGGGGGACAATCTGGGCCAGAAAGGCGCAGCTTGGC 1426
|||||
QY 20 aProProlaspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
|||||
Db 1425 ACCCCCTGATGCGCAGCATCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGCCCA 1366
|||||
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGlu-L 60
|||||
Db 1365 GGAGCTTTTTCAGGGCTCAGATTGGGCATGGCAGAGAGGCGCAGAGCCCTGGGGAGAA 1306
|||||
QY 60 ysAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuA 80
|||||
Db 1305 AAGCCGGCCAGCACAGCCCCCTGCGAGAGGAGCATGTGACCTGCGTACAGAGCATCTGG 1246
|||||
QY 80 spGluPheLeuGlnThr-TyrGlySerLeuIleProLeuSerThrAspGluValValGlu 99
|||||
Db 1245 ACGAATTCCTTCAACGGTATGGCAGCCTCATACCCCTCAGCACTGATGAGTAGTAGAG 1186
|||||
QY 100 LysLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeu 119
|||||
Db 1185 AAGCTGGAGGACATTTTCCAGCAGGAGTTTTCACCCCTTCCAGGAAGGCCCTGGTGTG 1126
|||||
QY 120 GlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgVal 139
|||||
Db 1125 CAGCTGATCCAGCTTACCAGCGGATGCCAGCAATGCCATGGTGGAGGGCTTCCGAGTG 1066
|||||
QY 140 AlaTyrLysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrp 159
|||||
Db 1065 GCTTATAAGCGGCAGCTGCTGACCATGGATGACTTTGGGGACCTGTGATGAGCAGAACTGG 1006
|||||
QY 160 LeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGlu-Ly 179
|||||
Db 1005 CTCATATGACCGGTGATGAACATGTATGAGACCTGGTTCATGCACACATCCCTCGAAAA 946
|||||
QY 179 sValHisPhePheAsnSerPhePheTyr-AspLysLeuArgThrLysGlyTyrAspGlyY 199
|||||
Db 945 GCTGCATTTCTCAATAGTTTCTCTATTGATAAACTCCGTACCAAGGGTTATGATGGG 886
|||||
QY 199 aLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleH 219
|||||
Db 885 TGAAGAGGTGGACAAAAACGTGGACATCTTCAATAAGAGGCTACTGCTTAATCCCATCC 826
|||||
QY 219 isLeuGluValHisTrpSerLeuIleSerValAspValArgArgArgThrIleThrTrp 239
|||||
Db 825 ACCTGGAGGTGCATTGGTCCCTCATCTCTGTTGATGTGAGGCGACGCCACCATCATTT 766
|||||
QY 239 heAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnA 259
|||||
Db 765 TTGACTCGCAGCGTACCCTAAACCGCGCTGCCCTAAAGCATATTGCAAGTATCTACAGG 706
|||||
QY 259 isGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysM 279
|||||
Db 705 CAGAGCGCGTGAAGAAAGACCGACTGGATTCCACCCAGGGCTGGAAAGGTACTTCAAAA 646
|||||
QY 279 etAsnValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysL 299
|||||
Db 645 TGAATGTGGCCAGGCAAGATAATGACAGTGACTGTGGTCTTTTGTGTGAGTACTGCA 586
|||||
QY 299 ysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgA 319
|||||

Db 585 AGCATCTGGCCCTCTCTCAGCCATTACAGTTCACCCAGCAGGACATGCCCAAACTTCGTC 526

Qy 319 rgGlnIleTyrLysGluLeuCysHisCysLysLeuThrVal 332

Db 525 GGCAGATCTACAGAGCTGTCTACATGCAAACTACTGTG 485

RESULT 7

AAI60212/c

ID AAI60212 standard; cDNA; 1804 BP.

XX AAI60212;

AC AAI60212;

XX 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 4201.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX Homo sapiens.

OS WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

DR P-PSDB; AAM41056.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 4201; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemia and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed

CC specification.

XX Sequence 1804 BP; 431 A; 477 C; 468 G; 428 T; 0 other;

SQ

Alignment Scores:		Pred. No.:	1.4e-187	Length:	1804
Score:		1699.00		Matches:	331
Percent Similarity:		98.22%		Conservative:	0
Best Local Similarity:		95.88%		Mismatches:	1
Query Match:		22		Indels:	5
DB:				Gaps:	0
US-09-848-852A-3 (1-332) x AAI60212 (1-1804)					
Qy	1	MetTyrSerAlaGlnArgPheTyrGlyThrIleTrrpAlaArgGlyAlaHis-LeuAl	20		
Db	1485	ATGTACTCTGCCAACGGTTTGGGGGACAATCTGGGCCCAAGAGGGGAGCTGTGC	1426		
Qy	20	aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl	40		
Db	1425	ACCCCTTGATGGCAGCATCTCATCAGCAATGTGTGCAGCATCGGGACCATTGGGCCCA	1366		
Qy	40	nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGlu-L	60		
Db	1365	GGAGCTTTTTCAGGGCTCAGATTGGGCATGGCAGAAGAGGAGAGGCTGGGGAGAA	1306		
Qy	60	ysAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuA	80		
Db	1305	AAGCGGCCAGCACAGCCCCCTCGAGAGGAGCATGTGACCTGCTACAGAGCATCTTG	1246		
Qy	80	spGluPheLeuGlnThr-TyrGlySerLeuIleProLeuSerThrAspGluValValGlu	99		
Db	1245	ACGAATTCCTTCAACGGTATGGCAGCCTCATACCCCTCAGCACATGATAGGTAGTAG	1186		
Qy	100	LysLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeu	119		
Db	1185	AAGCTGGAGGACATTTCCAGAGGAGTGTTCACCCCTCCAGGAAGGCCCTGGTGTG	1126		
Qy	120	GlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgVal	139		
Db	1125	CAGCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCTGTGAGGGGCTTCCGAGT	1066		
Qy	140	AlaTyrIlysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrp	159		
Db	1065	GCTTATAAGCGGCACGTGCTGACCATGGATGACTTGGGGACCTTGTATGACAGAACTGG	1006		
Qy	160	LeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGlu-Ly	179		
Db	1005	CTCAATGACCAGGCTGATGAACATGTATGGAGACCTGGTCATGGACACAGTCCCTGAAA	946		
Qy	179	sValHisPheAsnSerPhePheTyr-AspLysLeuArgThrLysGlyTyrAspGlyV	199		
Db	945	GGTGCATTTCTTCAATAGTTTCTTCTATTGATAAATCCGTACCAAGGGTTATGATGGG	886		
Qy	199	allysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleH	219		
Db	885	TGAAAAGGTGGACCAAAACGTGGACATCTTCAATAGGAGCTACTGCTAATCCCCATCC	826		
Qy	219	isLeuGluValHisTrrpSerLeuIleSerValAspValArgArgThrIleThrTyrP	239		
Db	825	ACCTGGAGGTGCATTGGTCCCTCATCTCTGTGATGTGAGCGCAGCCACCATCACTATT	766		
Qy	239	heAspSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrLeuGlnA	259		
Db	765	TTGACTCGCAGCGTACCCTAAACCGCGCTGCCCTAAGCATATTTGCCAAGTATCTACAGG	706		
Qy	259	laGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrrpLysGlyTyrPheLysM	279		
Db	705	CAGAGCGGTAAAGAAGACCGGACTGGATTTCACCAGGGCTGGAAAGGTATTCTCAAAA	646		
Qy	279	etAsnValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysL	299		
Db	645	TGAATGGCCAGGACAGATAATGACAGTCACTGTGTGTGCTTTGTGTGTGCTACTGCA	586		
Qy	299	ysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgA	319		
Db	585	AGCATCTGGCCCTGTCTCAGCCATTTCAGCTTCCAGGAGGAGGACATGCCCAAACTTCGTC	526		

```
QY 319 rgGlnIleTyrLysGluLeuCyHisCysLysLeuThrVal 332
|||||
Db 525 GCAGATCTACAAAGAGCTGTCTACTGCAAACTCACGTGTG 485

RESULT 8
ABQ54216
ID ABQ54216 standard; cDNA: 1933 BP.
XX
XX AC ABQ54216;
XX
XX DT 22-AUG-2002 (first entry)
XX
XX DE Human ovarian antigen HOF0B27 cDNA, SEQ ID NO:96.
XX
XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200200677-A1.
XX
XX PD 03-JAN-2002.
XX
XX PF 07-JUN-2001; 2001WO-US18569.
XX
XX PR 07-JUN-2000; 2000US-209467P.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Birse CE, Rosen CA;
XX
XX DR WPI: 2002-147878/19.
DR P-PSDB: ABP41139.
XX
XX PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
XX PS Claim 1; SEQ ID No 96; 2922pp; English.
XX
XX CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
XX ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
XX encompasses polypeptides 90% identical and polynucleotides 95% identical
XX to the sequences of the invention. The invention additionally relates to
XX recombinant vectors and host cells comprising human ovarian antigen
XX polynucleotides, antibodies against human ovarian antigens, and the use
XX of ovarian antigen polynucleotides and polypeptides in diagnosing,
XX treating, prognosing or preventing various ovary and/or breast-related
XX disorders. Such conditions include ovarian cancer and breast cancer, and
XX metastatic tumours of ovarian or breast origin, reproductive system
XX disorders (e.g., infertility, disorders of pregnancy, anovulation,
XX polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
XX disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
XX shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
XX vaginitis), immune disorders (e.g., congenital and acquired
XX immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
XX blood-related disorders (e.g., anaemia), cardiovascular disorders,
XX respiratory disorders, neurological disorders, gastrointestinal disorders
XX and urinary system disorders. Ovarian antigen polypeptides and
XX polynucleotides may also be used in screening for compounds which
XX modulate ovarian antigen expression or activity. The polynucleotides may
XX further be used for gene therapy, chromosome mapping, in the
XX identification of individuals and in forensic analysis, and the
```

```
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1933 BP; 423 A; 541 C; 551 G; 418 T; 0 other;

Alignment Scores:
Score: 1.27e-184 Length: 1933
Pred. No.: 1674.00 Matches: 323
Percent Similarity: 97.60% Conservative: 2
Best Local Similarity: 97.00% Mismatches: 7
Query Match: 94.47% Indels: 2
DB: 24 Gaps: 0

US-09-848-852A-3 (1-332) x ABQ54216 (1-1933)
QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTTPAlaAtgArgGlyAla-HisLeuAl 20
Db 496 CTGTACTCTTCCCAATGGCTTTGGGGACCCCGGGCCAGAAAGGGAGCGGGTCTGCG 555
QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
Db 556 ACCCCCTGATGCCAGCATCTTCATCAGCAATGTGTGCAGCATCGGGACCATGTGGCCCA 615
QY 40 nGluLeupheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLyl 60
Db 616 GGAGCTATTTCAGGGCTCAGATCTGGGCACCCAGCAAGAGGAGCGCGGGGAG-AA 674
QY 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 675 AGCCGGCCAGCACACGCCCTCGGGAGAGCATGTGACCTCGGTGAGAGCATCTTGA 734
QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLyl 100
Db 735 TGAATTCTCTCAAACTTATGGCAGCCTCATCCCTCAGCACGAGGAGTAGTAGAA 794
QY 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuLyl 120
Db 795 ATTAGAGGACATTTTCCAGCAGAGTCTCTACACCTTCACAGGAAGGGCTGTGCTGCA 854
QY 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 855 GCTGATCCAGTCGTACCGAGGATGCCAGGCAACGCCATGTGTGAGGGGCTTCCGGGTGCG 914
QY 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpIle 160
Db 915 CTACAAGCGGCACGTGCTGACCATGGATGACCTGGCACCCTTATATGGACAACTGGCT 974
QY 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
Db 975 CAACGACCAAGGTGATGATCAACATGTAGCGACCTGTGTCATGGACACGGTCCCGGAAAGT 1034
QY 180 IHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
Db 1035 GCATTCTTCAACAGTTTCTCTACGATAAGCTCCGACCAAGGGTTCAGATGGAGTGA 1094
QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuIleProIleHisLe 220
Db 1095 AAGGTGGACCAAAACGTGGACATCTTCAATAAGGAGCTCTGCTTAATCCCATCCACCT 1154
QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrPheAs 240
Db 1155 GGAGGTGCACTGGTCCCTCATCTCTGTGACGTGAGCGGCGCACCATCACGTATTTCGA 1214
QY 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGl 260
Db 1215 CTCGAGCGGACCCCTAAACCGCGCTGCGCTAAGCATATGTCCAAGTATCTACAGCGAGA 1274
QY 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280
```


|||||
Db 1275 GGCAGTGAAGAAGACCGCGTGGATTTCACACGCGCTGGAAGGTTATTTCAAAATGAA 1334
Qy 280 nValAlaAArgGlnAsnAsnAspSerAspCysGlyAlaAPheValLeuGlnTyrCysLysHI 300
Db 1335 TGTGGCCAGGCAGCAATAATGACAGTGACTGCGGTGCTTTGTGTACAGTACTGCAAGCA 1394
Qy 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGI 320
Db 1395 CTTGGCCCTGTCACGCCATTACAGTTCACCCAGCAGGACATGCCCAAACTTCGCCGGCA 1454
Qy 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
Db 1455 GATCTACAAGGAGGTGTGTCACTGCAAACTCACTGTG 1491
RESULT 9
AA158426
ID AA158426 standard; cDNA; 2029 BP.
XX AC AA158426;
XX DT 22-OCT-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 629.
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX P-PSDB; AAM39270.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX Claim 1; SEQ ID NO 629; 10078pp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
..

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX SQ Sequence 2029 BP; 460 A; 560 C; 553 G; 456 T; 0 other;
Alignment Scores:
Pred. No.: 6,39e-181 Length: 2029
Score: 1642.50 Matches: 314
Percent Similarity: 94.29% Conservative: 0
Best Local Similarity: 94.29% Mismatches: 1
Query Match: 92.69% Indels: 18
DB: 22 Gaps: 1
US-09-848-852A-3 (1-332) x AA158426 (1-2029)
Qy 1 MetTyrSerAlaGlnArgPheTyrGlyThrIleTyrAlaArgArgGlyAlaHis-LeuAl 20
Db 634 ATGTACTCTGCCCAACGGTTTTGGGGGACAACTCTGGGCCAGAGGGGAGCGAGCTTGGC 693
Qy 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGI 40
Db 694 ACCCCCTGATGCCAGCATCCTCATCAGCAATGTGTGCAGCATCGGGGACCATGTGCCCA 753
Qy 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLy 60
Db 754 GGAGCTTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGAGAGCCCTGGGGAGAA 813
Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 814 AGCCGCCACAGCAGCCCCCTGGGAGAGGAGCATGTGACCTGGTACAGAGCATCTTGA 873
Qy 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100
Db 874 CGAATTCTCTCAAACGTATGGCAGCCTCATACCCCTCAGCAGCTGATGAGGTAGTAGAA 933
Qy 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGI 120
Db 934 GCTGGAGGACATTTTCCAGCAGAGAGTTTTCCACCCCTTCCAGGAGGGGCTGGTGTGA 993
Qy 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 994 GCTGATCCAGTCTTACCACGGATGCCAGCAATGCCATGCTGAGGGGCTTCGAGTGGC 1053
Qy 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160
Db 1054 TTATAAGCGGCACGTGCTGACCATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCT 1113
Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
Db 1114 CAATGACCAAGGTGATGAACATGTATGGAGACCTGTGTCATGGACACAGTCCCTGAAAGGT 1173
Qy 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
Db 1174 GCATTTCTTCATAGTTTCTTCTATGATGATAAACTCCCTACCAAGGGTTATGATGGGGTGA 1233
Qy 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220
Db 1234 AAGTGGACCAAAACGTTGGACATCTTCAATAAGGAGCTACTGCTAATCCCCATCCACCT 1293
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgArgThrIleThrTyrPheAs 240
Db 1294 GGAGGTGCATTGGTCCCTCATCTCTGTGATGTGAGGGCAGCACCACCATCACCTATTTGA 1353
Qy 240 pSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGI 260
Db 1354 CTCGGACGCTACCTTAACCCCGCGCTGCCCTAAGCATATTTGCCAAGTATCTACAGGCAGA 1413
Qy 260 uAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAs 280
|||||

Db 1414 GCGGTAAGAAAGACCGACTGGATTTCACACAGGGCTGGAAGGTTACTTCAAAATG-- 1471
QY 280 nValAlaArgGlnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysIysShi 300
Db 1472 -----TACTGCAAGCA 1482
QY 300 sLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArg1 320
Db 1483 TCTGGCCCTGTCTCAGCCATTTCAGCTTCACCCAGCAGGACATGCCCAAACITTCGTCGGCA 1542
QY 320 nIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
Db 1543 GATCTACAAAGGAGCTGTCTCACTGCAAACTCACTGTG 1579
RESULT 10
AAH13708
ID AAH13708 standard; cDNA; 1578 BP.
XX AAH13708;
AC
XX
XX
XX 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:10590.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8; SEQ ID 10590; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1578 BP; 339 A; 372 C; 439 G; 428 T; 0 other;
Alignment Scores: 3.07e-85 Length: 1578
Pred. No.: 820.00 Matches: 186
Score: 50.00% Conservative: 10
Percent Similarity: 47.45% Mismatches: 17
Best Local Similarity: 46.28% Indels: 179
Query Match: 22 Gaps: 4
US-09-848-852A-3 (1-332) x AAH13708 (1-1578)
QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHis-LeuAl 20
Db 274 ATGTACTCTGCCCAACGGTTTTGGGGGACAATCTCTGGCCAGAAAGGGAGCGCGCTTGGC 333
QY 20 aProProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGl 40
Db 334 ACCCCTGATGCCAGCATCTCATCAGCAATGTGTGACATCGGGGACCATGTGGCCCA 393
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLyl 60
Db 394 GGAGCTTTTCAGGCTCAGATTTGGGCATGCCAAGAGGAGAGAGGCTTGGGAGAA 453
QY 60 sAlaGlyGlnHisSerProLeuArgGluHisValThrCysValGlnSerIleLeuAs 80
Db 454 AGCCGGCCAGCAGACGCCCTCGAGAGAGCATGTACCTCGGTACAGAGCATCTTGA 513
QY 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLyl 100
Db 514 CGAATTCCTTCAACGATATGGCAGCCTCATACCCCTCAGCACTGATGAGGTAGTAGAA 573
QY 100 sLeuGluAspIlePheGlnGlnPheSerThrProSer----- 113
Db 574 GCTGGAGGACATTTTCCAGCAGAGAGTTCACCCCTTCCAGGTCGAGGCTTGAAGCCCT 633
QY 113 ----- 113
Db 634 CCTTGAAGAAGGGCTGGGGCTTGGGGATGTGGAGAGAATACTGTCGCTTTTCTTCCA 693
QY 113 ----- 113
Db 694 TAGGCTCTAGTTGGGGAGAGGAAAGCTAGAGCTGAAGGGGAGGAGACTCTGCAGGAGGTG 753
QY 113 ----- 113
Db 754 CCAATCTTGGAAAGCTGATGGGAGAGTCTTTACCTGGGACCCCTGAATGTTCTACCTGA 813
QY 113 ----- 113
Db 814 GTAGTCATGTTTATCTTCTTGGGGAGTGGGGCTTTTCGAGGTCTCTCAGAAGAGCACCATCA 873
QY 113 ----- 113
Db 874 TGAGCCAGAAAAAAGGAGTCACTAGATTGTTATTCAGGGAGTAGTAGTATTTCTG 933
QY 114 -----ArgLysGlyLe 117
Db 934 TGTGCCCCAGCTGCATCATCTTTTGTGTGACTCCACCTTGGCCCTTACTCAGGAAGGGCCT 993
QY 117 uValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPh 137
Db 994 GGTGTTGCAGCTGATCCAGTCTTACCAGGGATGCCAGCAATGCATGCTGTGAGGGCTT 1053
QY 137 eArgValAlaTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGl 157
Db 1054 CCGAGTGGCTTATAAGCGGCACGTGCTGACCATGATGATCTTGGGGACCTTGTATGGACA 1113
QY 157 nAsnTrpLeuAsnAsp----- 162

```
|||||
Db 1114 GAACGGCTCAGTGACCAAGGTGAGAAAGGTGAGAAACAGGGCTGAGAGGGATTTCAGG 1173
QY 162 ----- 162
Db 1174 GAGCAGGGTGTCTGGGGCCCTCTGCATCGGGGGAGCCCTGTACCCATGCCGACCCCTCCAT 1233
QY 163 -----GlnValMetAsnMetTyrGlyAspLeuValMetAs 174
Db 1234 GGCAAGCTGCCTCCATCTTCTCCCCAGGTGATGAACATGTATGGACCTGTGTCATGGA 1293
QY 174 pThrValProGluLysVal-----HisPhePheAsnSe 185
Db 1294 CACAGTCCCTGAAAGGTAGGCCCAACACAGATAGGTAGTACCCAGAGGAACCTCTGCAG 1353
QY 185 rPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTrpThrLysAs 205
Db 1354 TTTT-----AAGCAGCTTTTAAAGCTCCCTTTCATCTTTCATTTTACACAGAGAG 1404
QY 205 nValAspIlePheAsnLysGluLeuLeuLeuIle 216
Db 1405 GGTCTCTGTTTCAGGAGAGAGGTGGTAGTGGA 1438
RESULT 11
AAI94288/c
ID AAI94288 standard; cDNA; 870 BP.
XX
AC AAI94288;
XX
DT 13-NOV-2001 (first entry)
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 363.
XX
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO200166719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP01629.
XX
PR 07-MAR-2000; 2000JP-0159195.
XX
PA (CHIB-) CHIBA PREFECTURE.
PA (HISM) HISAMITSU PHARM CO LTD.
XX
PI Nakagawara A;
XX
DR WPI; 2001-565584/63.
XX
PT Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents
XX
PS Claim 1; Page 308-309; 2979pp; Japanese.
XX
CC The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.
XX
SQ Sequence 870 BP; 217 A; 229 C; 190 G; 211 T; 23 other;
Alignment Scores:
Pred. No.: 2e-81 Length: 870
Score: 784.00 Matches: 174
Percent Similarity: 80.70% Conservative: 10
Best Local Similarity: 76.32% Mismatches: 29
```

```
Query Match: 44.24% Indels: 17
DB: 22 Gaps: 2
US-09-848-852A-3 (1-332) x AAI94288 (1-870)
QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIle-TripAlaArgArgGlyAlaHisLeuAl 20
Db 846 ATGAATTTTGGCCCAACGGTTTGGGGGACAATTTTGGCCCAAAAGGGAGCGCAGNTTGC 787
QY 20 aProPro-aspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG 40
Db 786 ACCCCCTTGATGCCACGATCCTCATCAGCAATGTGTGCAGCATCGGGGACCATNTGCCCC 727
QY 40 lncLysLeuPheGlnGlySerAspLeuGluAlaGluAlaGluArgProGlyGluL 60
Db 726 AGGA-CTTTTTCAGGNTTCAGATTTCGGGCATGCA-CANGAGCGCAGAGAGCGCTGGGGAGA 669
QY 60 ysaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIle-Leu 79
Db 668 AAGCCGGCCAGCAGCACGCCCTTGCAGAGAGGAGCATGTGACTTGCCTACAGAGCATCTTG 609
QY 80 AspGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGlu 99
Db 608 GACGAATTCCTTCAAACGATATGCAGCNCATACCCNTCAGCAGCATGATGAGGTAGTAGAG 549
QY 100 LysLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeu 119
Db 548 AAGCTGGAGAGCATTTTCCAGCAGGAGTTTTCACCCCTTCCAGGAGGCCCTGGTGTG 489
QY 120 GlnLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgVal 139
Db 488 CAGCTNATCCAGTCTTTACCAGCGGATGCCAGGCAATGCCATGGTGGAGGGCTTCCGAGTG 429
QY 140 AlaTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrp 159
Db 428 GCTTATAAGCGGCACGCTGTGACCATGGATGACCTTGGGACCTTGTATGGACAGAACTGG 369
QY 160 LeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLys 179
Db 368 CTCATCACCAGGTGATGAACATGTATGGAGACCTGGTGCATGGACACAGTCCCTGAAAG 309
QY 180 Val-----HisPhePheAsnSerPhePheTyrAspLys 190
Db 308 GTAGGCCCAACACAGATAGTACAGTACCCAGAGAACTTCTGCAGTTT-----AAG 258
QY 191 LeuArgThrLysGlyTyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsn 210
Db 257 CAGCTTTTAAAGCTCCTTTTCATCTCTTTTATTTTACACAGAGAGGGTCTCTGTTTCAGGG 198
QY 211 LysGluLeuLeuIle 216
Db 197 AGAGAGGTGGTAGTGGA 180
RESULT 12
AAH99558
ID AAH99558 standard; cDNA; 1938 BP.
XX
AC AAH99558;
XX
DT 16-OCT-2001 (first entry)
DE Human protein encoding cDNA sequence SEQ ID NO:393.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antinflammatory; antirheumatic; antiallergic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
KW antiagregant; haemostatic; vulnary; antileuk; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
```

KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder; ss.

XX Homo sapiens.

OS WO200153455-A2.

PN 26-JUL-2001.

PD 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

XX WPI, 2001-457603/49.

DR P-PSDB; AAM25617.

XX Isolated human polynucleotides encoding polypeptides, useful for the
 treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX Claim 1; Page 494; 1217pp; English.

CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
 CC antiulcer; osteopathic; dermatological; anti allergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.

XX Sequence 1938 BP; 565 A; 366 C; 425 G; 581 T; 1 other;

Alignment Scores:

Pred. No.: 1.36e-77 Length: 1938
 Score: 755.50 Matches: 148
 Percent Similarity: 73.19% Conservative: 54
 Best Local Similarity: 53.62% Mismatches: 65
 Query Match: 42.64% Indels: 9
 DB: 22 Gaps: 4

US-09-848-852A-3 (1-332) x AAH99558 (1-1938)

QY 58 GlyGluLysAlaGlyClnHisSerProLeuArgGluGluHisValThr---CysValGln 76

DB 3 GGAAGACACAGTCAGAAAGCCCTCTCTGGTGATGATGAACAGCTGTCAGTCTGCTTCT 62

QY 77 SerIleLeuAspGluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGlu 96

DB 63 GGATTCCTAGATGAGTTATGAAGAAGTATGGCAGTTTGGTTCCACTCAGTGAAGAA 122

QY 97 ValValGluLysLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGly 116

DB 123 GTCTTGAAGATATAAAGATGCTTTAATGAAGACTTTTCT-----AATAGAAACCA 176

QY 117 LeuValLeuGlnLeuIleGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGly 136

DB 177 TTTATCAATAGGGAATAACAACTATCGGCCAGACATCAAAAATGT-----AAC 227

QY 137 PheArgValAlaIaTyrglyLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrgly 156

DB 228 TTCCGTATCTTCTATAATAAACACATGCTGGATATGGAGCAGCTCTGGATGTT 287

QY 157 GlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrVal 176

DB 288 CAGAACTGGCTCAATGACCAAGCTCATTAATATGATGTGAGCTGATATGATGATGATG 347

QY 177 ProGluLysValHisPhePheAsnSerPhePheTyrglyAspLeuArgThrLysGlyTy 196

DB 348 CCAGACAAAGTTTCATCTTCAACAGCTTTTTCATAGACAGCTGGTAACCAAGGATAT 407

QY 197 AspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeu 216

DB 408 AATGGAGTAAAGATGGACTAAAGGTGGATTTGTTTAAAGAGAGTCTTCTGTGAT 467

QY 217 ProIleHisLeuGluValHisTrpSerLeuIleSerValAspValArgArgThrIle 236

DB 468 CCTATTCACTGGAAGTCCACTGGTCTCTCATTTACTGTGACACTCTCTAATCGAATTAT 527

QY 237 ThrTyrglyAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAlaLysTy 256

DB 528 TCATTATTATGATCCCAAGGCATTCAATTTAAGTTTGTGTAGAGAAATATAAGAAGTAT 587

QY 257 LeuGlnAlaGluAlaValLysLysAspArg-LeuAspPheHisGlnGlyTrpLysGlyTy 276

DB 588 TTCTGCTGACGAGCCAGACAGAGAAATAATAGACTGTAATCTT---CAGGGTTGCAGACTGC 644

QY 276 rPheLysMetAsnValAlaArgGlnAsnAspSerAspCysGlyAlaPheValLeuGlu 296

DB 645 TGTACGAGAGTCTATTCCACACAGACAAAACACAGTACTGTGGAGTCTTTGTGCTCCA 704

QY 296 nTyrglyLysHisLeuAla-LeuSerGlnProPheSerPheThrGlnGlnAspMetPro 316

DB 705 GTACTGCAAGTCCCTCCGCTTTAGAGCAGCCTTTCCAGTTTTCACAAAGAAGACATGCC 764

QY 316 ysLeuArgArgGlnIleTyrglyLysGluLeuCysHisCysLysLeu 330

DB 765 GAGTCCGGAAGAGGATTACAAAGAGGCTATGTGAGTCCCGGCTC 808

RESULT 13

AA25529/c

ID AA25529 standard; cDNA; 503 BP.

XX AA25529;

XX 07-NOV-2001 (first entry)

XX Human ovarian PCR-subtracted cDNA library clone #1614.

XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;
 KW gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;
 KW primer; probe.

XX Homo sapiens.

XX WO200157207-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US03733.

XX 04-FEB-2000; 2000US-0180403.

XX 28-MAR-2000; 2000US-0192745.

XX (CORI-) CORIXA CORP.

Db 122 ACCATCACCTATTTCGACTCGCAGCGTACCCTAAACCGCGCTGCCCTAAGCATATTGCC 181
 QY 255 LysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLys 274
 Db 182 AAGTATCTACAGCGCAGCGCGTAAAGAAGACCAGCTGGATTTCACCCAGGCGCTGGAAA 241
 QY 275 GlyTyrPheLysMetAsnValAlaArgGlnAsnAspSerAspCysGlyAlaPheVal 294
 Db 242 GCTTACTTCAAAATG----- 256
 QY 295 LeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMet 314
 Db 257 -----TACTGCAAGCATCTGCCCTGTCTCAGCCATTTCAGCTTCACCCAGGACATG 310
 QY 315 ProLysLeuArgGlnIleTyrLysGluLeuCysHisCysLysLeu 330
 Db 311 CCCAACTTCGTGCGCAGATCTACAAAGGAGCTGTCTACTGCAAACTN 358
 RESULT 15
 AA193958/c
 ID AA193958 standard; cDNA; 820 BP.
 AC AA193958;
 XX
 DT 13-NOV-2001 (first entry)
 XX
 DE Human neuroblastoma expressed polynucleotide SEQ ID NO 33.
 XX
 KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200166719-A1.
 XX
 PD 13-SEP-2001.
 XX
 PF 02-MAR-2001; 2001WO-JP01629.
 XX
 PR 07-MAR-2000; 2000JP-0159195.
 XX
 PA (CHIB-) CHIBA PREFECTURE.
 PA (HISM) HISAMITSU PHARM CO LTD.
 XX
 PI Nakagawara A;
 XX
 DR WPI; 2001-565584/63.
 XX
 PT Nucleic acids originating in gene expressed in human neuroblastoma,
 PT useful as probe or primer in diagnosing prognosis of human
 PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
 PT for anti-cancer agents -
 XX
 PS Claim 1; Page 59; 2979pp; Japanese.
 XX
 CC The invention relates to novel genes (AA193926-AA197963) expressed in
 CC human neuroblastoma. The nucleic acids are applicable as a probe or
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
 CC susceptibility indicators or tumour markers for anti-cancer agents. The
 CC gene information for diagnosing prognosis is related to factors similar
 CC to that for N-myc and TrkA genes.
 XX
 SQ Sequence 820 BP; 224 A; 223 C; 170 G; 191 T; 12 other;

Alignment Scores:

Pred. No.:	1.18e-61	Length:	820
Score:	614.00	Matches:	149
Percent Similarity:	72.07%	Conservative:	11
Best Local Similarity:	67.12%	Mismatches:	46
Query Match:	34.65%	Indels:	19
DB:	22	Gaps:	5

US-09-848-852A-3 (1-332) x AA193958 (1-820)

QY 5 GlnArgPheTrpGlyThrIleTrpAlaArgArgGlyAlaHisLeuAlaProProAspAla 24
 Db 784 CAAACGTTTTGGGGAA-ATTTGGCCCNAGGGGACGCGAGTTTGACCCCTTG---ATC 729
 QY 25 SerIleLeuIleSerAsnValCysSerIleGlyAspHisValalaglnGluLeuPheGln 44
 Db 728 CAGATCCTTATCAGCAATNTTTTCAGCATCGGGGACA--TNTGCCACGAGAGNT-TTTCAG 673
 QY 45 GlySerAspLeuGlyMetAlaGluGluAlaGluArgProGlyGluLysAlaGlyGlnHis 64
 Db 672 GGTTCAGATT--GGCATGCAGAGGACGAGAGCCCTGGGGAGAA---GCCGGCCAGAAC 617
 QY 65 SerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAspGluPheLeuGln 84
 Db 616 AGCCCTTNGAGAGGAGCATGTGATTCGGAACAGAGCATTTTGACCANITCCTTCAA 557
 QY 85 -ThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLysLeuGlnAsp 104
 Db 556 AACGTATGCGACGNTCATCCCTTCAGCACTGATGAGGTAGTAGAAGTTGGAGGACAT 497
 QY 104 ePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSe 124
 Db 496 TTTCCAGCAGGAGTTTTCACCCCTTCCAGGAAGGCTTGGTTCGACCTNATCCAGTC 437
 QY 124 rTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAlaLysLysArgHi 144
 Db 436 TTACACGCGGATGCCAGCAATGCCATGCTGAGGGCTTCCAGTGGCTTATAAGCGGCA 377
 QY 144 sValLeuThrMetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnVa 164
 Db 376 CGTGCTGACCATGGATGACTTGGGACCTTGTATGGACAGAACTGGCTCAATGACCGGT 317
 QY 164 lMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVal----- 180
 Db 316 GATGAACATGATGAGAGACCTGGTCATGACACACGTCCTGAAAGGTAGGCCCAACACAG 257
 QY 181 -----HisPheAsnSerPhePheTyrAspLysLeuArgThrLysGln 195
 Db 256 ATAGGTCAGTACCAGAGGAACCTTCTGCAGTTTT-----AAGCAGCTTTTAAAGCT 206
 QY 195 yTyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLe 215
 Db 205 CCTTTCATCTCTTTCATTTTACACAGAGGCTCTCTGTTTCAGGAGAGAGGTGTGTAGT 146
 QY 215 uIle 216
 Db 145 GGT 142

Search completed: December 16, 2002, 21:27:46

Job time : 313 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2002, 14:21:29 ; Search time 5036 Seconds
(without alignments)
11505.895 Million cell updates/sec

Title: US-09-848-852A-4
Perfect score: 1991
Sequence: 1 GCCTCCCTGTCCTCCGACCC.....TGCTCTGGTCAATAAGATC 1991

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_vl.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1928.6	96.9	2269	9	HSB805332	AL834294 Homo sapi
2	1925.4	96.7	2258	9	AF199459	AF199459 Homo sapi
3	1907.8	95.8	2224	9	AY008763	AY008763 Homo sapi
4	1843.4	92.6	2206	6	AX081143	AX081143 Sequence
5	1692.4	85.0	1782	9	HSB800198	AL050283 Homo sapi
6	1459.2	73.3	2174	10	AY008764	AY008764 Mus muscu
7	1279	64.2	1707	10	AF194031	AF194031 Mus muscu
8	534	26.8	17703	2	AC016876	AC016876 Homo sapi
9	530.4	26.6	164504	2	AC129071	AC129071 Pan trogl
10	522.4	26.2	173588	2	AC127470	AC127470 Papio cyn
11	486.4	24.4	5318	6	AX069229	AX069229 Sequence
12	486.4	24.4	6732	9	AF175325	AF175325 Homo sapi
13	477.4	24.0	164504	2	AC129071	AC129071 Pan trogl
14	474.4	23.8	161428	2	AC126925	AC126925 Canis fam
15	474.4	23.8	203281	2	AC126237	AC126237 Canis fam
16	469.6	23.6	139405	2	AC126239	AC126239 Felis cat
17	464.8	23.3	177555	2	AC130192	AC130192 Sus scrof
18	460	23.1	153553	2	AC126921	AC126921 Bos tauru
19	442	22.2	503	6	AX209870	AX209870 Sequence
20	427.6	21.5	184026	2	AC098923	AC098923 Rattus no
21	426	21.4	234182	10	AL603707	AL603707 Mouse DNA
22	344.6	17.3	1578	9	AK000923	AK000923 Homo sapi
23	333.4	16.7	2007	9	AB060892	AB060892 Macaca fa
24	301.6	15.1	138792	2	AC119115	AC119115 Rattus no
25	277.4	13.9	209237	2	AL806532	AL806532 Mus muscu
26	260.4	13.1	257644	2	AC099294	AC099294 Rattus no
27	254.6	12.8	2673	9	AB074445	AB074445 Macaca fa
28	253.8	12.7	251835	2	AC099436	AC099436 Rattus no
29	251.4	12.6	1830	9	BC008589	BC008589 Homo sapi
30	251.4	12.6	2793	9	BC030705	BC030705 Homo sapi
31	230.6	11.6	76754	2	AC097861	AC097861 Rattus no
32	183.4	9.2	333300	2	AC125091	AC125091 Mus muscu
33	182.4	9.2	138792	2	AC119115	AC119115 Rattus no
34	176	8.8	1424	10	MUSELF4A1	M22873 Mus musculu
35	160.2	8.0	182718	2	AC103148	AC103148 Rattus no
36	143.8	7.2	209237	2	AL806532	AL806532 Mus muscu
37	101.4	5.1	2903	10	BC023129	BC023129 Sequence
38	98.4	4.9	2512	6	AX081135	AX081135 Sequence
39	96	4.8	1932	9	AF149770	AF149770 Homo sapi
40	83.8	4.2	763	9	HSA343855	AJ343855 Homo sapi
41	83.8	4.2	964	9	HSA343584	AJ343584 Homo sapi
42	83.2	4.2	2142	10	AF368904	AF368904 Mus muscu
43	83.2	4.2	3109	10	BC031652	BC031652 Mus muscu
44	81.6	4.1	1808	10	AF260129	AF260129 Rattus no
45						

ALIGNMENTS

RESULT 1
LOCUS HSB805332 2269 bp mRNA linear PRI 12-JUL-2002
DEFINITION Homo sapiens mRNA; cDNA DKFP762A152 (from clone DKFP762A152).
ACCESSION AL834294
VERSION AL834294.1 GI:21739883
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2269)
AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY

Db 1642 ATGTGCCAGGCGAATAATGACAGTACTGTGGTCTTTTGTGTGCAGTACTGCAAGC 1701
QY 1440 ATCTGCCCTCTCTCAGCCATTTCAGCTTCACCCAGCAGGACATGCCAAATCTCGTCGCG 1499
Db 1702 ATCTGCCCTCTCTCAGCCATTTCAGCTTCACCCAGCAGGACATGCCAAATCTCGTCGCG 1761
QY 1500 AGATCTACAAGGAGCTGTCTACTGCAAACTCACTGTGTGAGCCCTCGTACCCAGACCCC 1559
Db 1762 AGATCTACAAGGAGCTGTCTACTGCAAACTCACTGTGTGAGCCCTCGTACCCAGACCCC 1821
QY 1560 AAGCCCATAAATGGGAAGGAGACATGGAGTCCCTTCCCAAGAACTCCAGTTCTCTTC 1619
Db 1822 AAGCCCATAAATGGGAAGGAGACATGGAGTCCCTTCCCAAGAACTCCAGTTCTCTTC 1881
QY 1620 CTCCTCTGCTCTCTCCCACTCACTTCCCTTGTGTTTTCATATTTAAATGTTTCAATTC 1679
Db 1882 CTCCTCTGCTCTCTCCCACTCACTTCCCTTGTGTTTTCATATTTAAATGTTTCAATTC 1941
QY 1680 TGTATTTTCTTTCTTGAGAGAATACCTTGTTGATTTCTGATGTGCGAGGGGTGGCTACA 1739
Db 1942 TGTATTTTCTTTCTTGAGAGAATACCTTGTTGATTTCTGATGTGCGAGGGGTGGCTACA 2001
QY 1740 GAAAAGCCCTTCTCTCTCTGTCAGGGAGTGTGCCCCCTGTGGCCCTGGGTGAGCA 1799
Db 2002 GAAAAGCCCTTCTCTCTCTGTCAGGGAGTGTGCCCCCTGTGGCCCTGGGTGAGCA 2061
QY 1800 GTCATCTCCCTTCCCGTGCAGGAGCAGGAATCACTGCTGGGGTGTGGCGGA 1859
Db 2062 GTCATCTCCCTTCCCGTGCAGGAGCAGGAATCACTGCTGGGGTGTGGCGGA 2121
QY 1860 CAATAGGATCACTGCTGCGAGATCTTCAAACCTTTATATATATATATATATATATATAT 1919
Db 2122 CAATAGGATCACTGCTGCGAGATCTTCAAACCTTTATATATATATATATATATATATAT 2181
QY 1920 AT 1979
Db 2182 AT 2225
QY 1980 TCAATAAAG 1988
Db 2226 TCAATAAAG 2234

RESULT 2 AF199459 2258 bp mRNA linear PRI 31-OCT-2001
LOCUS Homo sapiens SUMO-1 specific protease 3 (SSP3) mRNA, complete cds.
DEFINITION AF199459
ACCESSION AF199459.1 GI:16550942
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Choi,S.J., Jeon,Y.J., Kim,K.I., Nishimori,S., Suzuki,T., Uchida,S.,
Shimbara,N., Tanaka,K. and Chung,C.H.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1999) Molecular Biology, Seoul National
University, Shillim-dong san 56-1, Seoul 151-742, Korea
FEATURES
source
1..2258
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2258
/gene="SSP3"
71..1795
/gene="SSP3"
/codon_start=1
/product="SUMO-1 specific protease 3"
/protein_id="AAL25652.1"
/db_xref="GI:16550943"
/translation="MKETIQGTGSGPPEPGIPPAYSSPRRLRWPPPPKRLKS

GGGFGDPGSGTTPARRLPVPRPSFDASAESEEEEEDEDEEEVAAWRLPPRR
SQLTSQRPSRPTHRKTCQRRRRAMRFLYSKSTSLTTHWKLWGRGRRRG
LAHPKHLSPQGGATPOVSPPCRFDPGRPPRLGLGLMAEDGVRGSPVPVSG
PMEDGRLWTPKSPLDPSGLSCTLPNGFGSGSGPESGSLAPDSAILLSNVCSI
GDVAAQELFQSGDLGMAEABRPEKAGQHSPLREHVTVCQSIIDFELQTVGSLIPL
STDVEVKLEDIFQDEFSPSRKGLVLQLOISYRMPGNAMVGRFVRYKRVHVTMDI
LGTLTGONLNDQVNMNYGDLVMDTPPEKVFHFFNSFYDKLRTKYGDKVKTWTKNDI
FNKELLIPILHLEVHWSLISDVMDRTITYFDSQRTLNRRCPKHLAKYLAQBAVKKDR
LDFHGKGYFKMNVARQNDSDCGAFVLYQCKHLALSQPFSEFTQDDMPKRLRQYKE
LCHKLVTV"
BASE COUNT 519 a 626 c 618 g 495 t
ORIGIN
Query Match 96.7%; Score 1925.4; DB 9; Length 2258;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1967; Conservative 0; Mismatches 1; Indels 21; Gaps 2;
QY 1 GCCTCCCTCTCCCGAACCTCTTTTGATGCCTACAGTGAAGAGGAGGAGAGAGG 60
Db 255 GCCTCCCTCTCCCGAACCTCTTTTGATGCCTACAGTGAAGAGGAGGAGAGAGG 314
QY 61 AGGAGAGGAGGATGAAGATGAAGAGGAGGAGTGGCAGCTTGGAGGCTGCCCCCAAGAT 120
Db 315 AGGAGAGGAGGATGAAGATGAAGAGGAGGAGTGGCAGCTTGGAGGCTGCCCCCAAGAT 374
QY 121 GGAGTCAGCTGGGAACCTCCCGAGGSCCGCCCTTCCCGCCCACTCATCGAAAAACCT 180
Db 375 GGAGTCAGCTGGGAACCTCCCGAGGSCCGCCCTTCCCGCCCACTCATCGAAAAACCT 434
QY 181 GCTCAGAGCGCCCGCGAGCCATGAGAGCCCTTCGGATGCTGTCTACTCAAAAAGCA 240
Db 435 GCTCAGAGCGCCCGCGAGCCATGAGAGCCCTTCGGATGCTGTCTACTCAAAAAGCA 494
QY 241 CTTGCTGATATTCACACTGGAAGCTTTGGGGGGCCACCGGGGGCGGGGGGGCTCG 300
Db 495 CTTGCTGATATTCACACTGGAAGCTTTGGGGGGCCACCGGGGGCGGGGGGGCTCG 554
QY 301 CACACCCCAAGAACCATCTTTTCCACCCAGAGGGGTGCGACCCACAGGTGCCATCCC 360
Db 555 CACACCCCAAGAACCATCTTTTCCACCCAGAGGGGTGCGACCCACAGGTGCCATCCC 614
QY 361 CTTGTTGCTGTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGGTGGGTGCTAGGTG 420
Db 615 CTTGTTGCTGTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGGTGGGTGCTAGGTG 674
QY 421 CTCATGCTCAGGATGGGTGAGAGGCTTCCACCACTGCTGGGGGGGGGGGGGGGGGG 480
Db 675 CTCATGCTCAGGATGGGTGAGAGGCTTCCACCACTGCTGGGGGGGGGGGGGGGGGG 734
QY 481 AGGAAGATGGATCAGTGGACTCCAAAGTCTCTCTGACCCCTGACTCGGGCTCTTTT 540
Db 735 AGGAAGATGGATCAGTGGACTCCAAAGTCTCTCTGACCCCTGACTCGGGCTCTTTT 794
QY 541 CATGTACTCTGCCAACCGTTTTTGGGGGACAATCTGGGGCAGAGGGGGAGCCCA-CTTGG 599
Db 795 CATGTACTCTGCCAACCGTTTTTGGGGGACAATCTGGGGCAGAGGGGGAGCGAGCTTG 854
QY 600 CACCCCTCATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGCCC 659
Db 855 CACCCCTCATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGCCC 914
QY 660 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGGAGGAGGAGGAGGAGGAG 719
Db 915 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGGAGGAGGAGGAGGAGGAGGAG 974
QY 720 AAGCCGGCCAGCACAGCCCCCTCGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTG 779
Db 975 AAGCCGGCCAGCACAGCCCCCTCGAGAGGAGCATGTGACCTCGGTACAGAGCATCTTG 1034
QY 780 AGCAATTCCTCAAACGTATGGCAGCCTCATACCCCTCAGCAGCATGTAGGTAGTAGAGA 839
Db 1035 AGCAATTCCTCAAACGTATGGCAGCCTCATACCCCTCAGCAGCATGTAGGTAGTAGAGA 1094

QY 840 AGCTGGAGCAGATTTTCCAGCAGAGTTCACCCCTTCCAGGAGGCTGTGTTC 899
Db 1095 AGCTGGAGGACATTTTCCAGCAGAGTTCACCCCTTCCAGGAGGCTGTGTTC 1154
QY 900 AGCTGATCCAGTCTTACAGGGATGCCAGGCAATGCCATGGTGAGGGCTTCGAGTGG 959
Db 1155 AGCTGATCCAGTCTTACAGGGATGCCAGGCAATGCCATGGTGAGGGCTTCGAGTGG 1214
QY 960 CTTATAGCGCAGCTGCTGACCATGATGATGGGACCTGTATGGACAGCACTGSC 1019
Db 1215 CTTATAGCGCAGCTGCTGACCATGATGATGGGACCTGTATGGACAGCACTGSC 1274
QY 1020 TCAATGACCGAGGTGATGAACATGATGAGACCTGCTGATGGACACAGTCCCTGAAAAGG 1079
Db 1275 TCAATGACCGAGGTGATGAACATGATGAGACCTGCTGATGGACACAGTCCCTGAAAAGG 1334
QY 1080 TGCATTTCTCAATAGTTTCTTCTATGATAAATCCCTACAGGGTATGATGGGGTGA 1139
Db 1335 TGCATTTCTTCAATAGTTTCTTCTATGATAAATCCCTACAGGGTATGATGGGGTGA 1394
QY 1140 AAGGTGGACCAAAAGCGTGACATCTTCAATAGGAGCTACTGCTAATCCCATCCACC 1199
Db 1395 AAGGTGGACCAAAAGCGTGACATCTTCAATAGGAGCTACTGCTAATCCCATCCACC 1454
QY 1200 TGGAGGTGCATTTGGTCCCTCATCTCTCTGATGTGAGCGCAGCACCATCACTATTTCG 1259
Db 1455 TGGAGGTGCATTTGGTCCCTCATCTCTCTGATGTGAGCGCAGCACCATCACTATTTCG 1514
QY 1260 ACTCGAGCGTACCTAAACGGCGCTGCCCTAAGCAATNTGCCAAGTATCTACAGGCAG 1319
Db 1515 ACTCGAGCGTACCTAAACGGCGCTGCCCTAAGCAATNTGCCAAGTATCTACAGGCAG 1574
QY 1320 AGCGGTAAAGAAAGCGAGTGGATTTCCACAGGCTGGAAGGTACTTCAAAATGA 1379
Db 1575 AGCGGTAAAGAAAGCGAGTGGATTTCCACAGGCTGGAAGGTACTTCAAAATGA 1634
QY 1380 ATGTGGCCAGCAGAAATATGACAGTGACTGTGTGCTTTTGTGTTGCAGTACTGCAAGC 1439
Db 1635 ATGTGGCCAGCAGAAATATGACAGTGACTGTGTGCTTTTGTGTTGCAGTACTGCAAGC 1694
QY 1440 ATCTGGCCCTGTCTACGCCATTCAGCTTACCAGCAGGACATGCCCAAATCTGCTGGC 1499
Db 1695 ATCTGGCCCTGTCTACGCCATTCAGCTTACCAGCAGGACATGCCCAAATCTGCTGGC 1754
QY 1500 AGATCTACAAGGAGCTGTCTACGTCAAACTCACTGTGAGCTCTGCTACCCAGACCCC 1559
Db 1755 AGATCTACAAGGAGCTGTCTACGTCAAACTCACTGTGAGCTCTGCTACCCAGACCCC 1814
QY 1560 AAGCCCAATAATGGGAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC 1619
Db 1815 AAGCCCAATAATGGGAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC 1874
QY 1620 CTCTCTTGCCTCTCCGACCTCACTTCCTTGGTTTTCATATTAATGTTTCAATTC 1679
Db 1875 CTCTCTTGCCTCTCCGACCTCACTTCCTTGGTTTTCATATTAATGTTTCAATTC 1934
QY 1680 TGTATTTTCTTTTGTGAGAGAACTCTGATTTCTGATTTCTGATGTCAGGSGGTGCTACA 1739
Db 1935 TGTATTTTCTTTTGTGAGAGAACTCTGATTTCTGATTTCTGATGTCAGGSGGTGCTACA 1994
QY 1740 GAAAGCCCTTCTTCTCTCTGTTGAGGGAGTGTGGCCCTGTGCGCTGGGTGGAGCA 1799
Db 1995 GAAAGCCCTTCTTCTCTCTGTTGAGGGAGTGTGGCCCTGTGCGCTGGGTGGAGCA 2054
QY 1800 GTCATCTCCCTTCCCTGTGAGGAGCAGGAAATCACTGCTGGGGTGGTGGCGGA 1859
Db 2055 GTCATCTCCCTTCCCTGTGAGGAGCAGGAAATCACTGCTGGGGTGGTGGCGGA 2114
QY 1860 CAATAGATCACTGCGTCCAGATCTTCAACCTTTTATATATATATATATATATATAT 1919
Db 2115 CAATAGATCACTGCGTCCAGATCTTCAACCTTTTATATATATATATATATATATATAT 2154
QY 1920 AT 1979

Db 2155 AT 2214
QY 1980 TCAATAAG 1988
Db 2215 TCAATAAG 2223
RESULT 3
AY008763 2224 bp mRNA linear PRI 21-NOV-2000
LOCUS Homo sapiens sentrin/SUMO-specific protease (SEN3) mRNA, complete.
DEFINITION cds.
ACCESSION AY008763
VERSION AY008763.1 GI:11245810
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2224)
AUTHORS Yeh,E.T., Gong,L. and Kamitani,T.
TITLE Ubiquitin-like proteins: new wines in new bottles
JOURNAL Gene 248 (1-2), 1-14 (2000)
MEDLINE 20267842
PUBMED 10806345
REFERENCE 2 (bases 1 to 2224)
AUTHORS Gong,L. and Yeh,E.T.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2000) Institute of Molecular Medicine, University of Texas-Houston, 2121 W. Holcombe Blvd. 928, Houston, TX 77030, USA
FEATURES
source Location/Qualifiers
1..2224
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p13"
1..2224
/gene="SEN3"
52..1776
/gene="SEN3"
/codon_start=1
/product="sentrin/SUMO-specific protease"
/protein_id="AAG33252.1"
/db_xref="GI:11245811"
/translation="MKETIQGTSNGPEPPGIPPAYSSPRDGLRWPPPKPLKLS
GGFGPDGSGTTPVTRRLPAPRPSFDASASEEEEEDEEEDEEEVAARLPRW
SOLGTSQRPALRPSHRKTCSQRRRRAMRAFQMLLYSKSTSLTHWKLWGRHRRRG
LAHPKNHLSPOCGATPOVSPCCRFDSPPRGLGLGALMAEDVGRGPPVPSG
PMWEDGLRWTPKSPLDPSGLLSCTLPNGFGSGSGPEGRSLAPDASILISNVC
GDHVAQELFGSDLGMAEAEAPRPEKAGHSPLEHVTVCQSILDEFLTQYGLSLPL
STDEVVEKLEIFQEFSTPSRRGLVLQISQIRMPGNAMVGRVATKRHLVTMD
LGTLGQNNLNDVNMVMDLYMDVPEKHFNFYFDKRTKGTGVDYKRWTKNVDI
FNKEILLPIHLEVHWSLISVDVRRRTITYFDSQRTLNRRCPKHIKAYLQAEAVKDR
LDFHOGWKGYFKMNVARQNDSDCGAFVLQYCKHLALSQPFSTOODMPKLRQIYE
LCHCKLTV"
BASE COUNT 502 a 615 c 612 g 495 t
ORIGIN
Query Match 95.8%; Score 1907.8; DB 9; Length 2224;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1956; Conservative 0; Mismatches 12; Indels 21; Gaps 2;
QY 1 GCCTCCTGTCCCGGACCCCTCTTTTGATGCGCTCAGCAGTGAAGAGGAGGAAGAAGG 60
Db 236 GCCTCCTGTCCCGGACCCCTCTTTTGATGCGCTCAGCAGTGAAGAGGAGGAAGAAGG 295
QY 61 AGCAGGAGGAGGATGAAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
Db 296 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 355
QY 121 GGAGTCAGCTGGGAACCTCCAGCGCGCCCTTCCCGCGCCCACTCATCATGAAAAAACCT 180

Db 356 |||||GGAGTCAGTGGGAACCTCCACAGCGCGCCGCGCTCTCCGACCCCTCATAGAAAACCT 415
QY 181 GCTCAGCGCGCGCGCGAGCATGAGAGCCTTCGGATGCTGCTCTACTCAAAAAGCA 240
Db 416 GCTCAGCGCGCGCGCGAGCCATGAGAGCCTTCAGATGCTGCTCTACTCAAAAAGCA 475
QY 241 CCTCGCTGACATTCACATCGAAAGCTTTGGGGGCGCCACCGGGCGCGCGCGCGCGCTCG 300
Db 476 CCTCGCTGACATTCACATCGAAAGCTTTGGGGGCGCCACCGGGCGCGCGCGCGCGCTCG 535
QY 301 CACACCCCAAGAACATCTTTACCCACAGCAAGGGGCTGCGACGCCACAGGTGCGCATCC 360
Db 536 CACACCCCAAGAACATCTTTACCCACAGCAAGGGGCTGCGACGCCACAGGTGCGCATCC 595
QY 361 CCTGTTGTGCTTTGACTCCCGCGGGGCGCCACCTCCACCGCGGCTGGGTCTGCTAGGTG 420
Db 596 CCTGTTGTGCTTTGACTCCCGCGGGGCGCCACCTCCACCGCGGCTGGGTCTGCTAGGTG 655
QY 421 CTCTCATGGCTGAGATGGGCTGAGAGGCTCTCCACCAAGTCCCTCTGGGCGCGCGCATGG 480
Db 656 CTCTCATGGCTGAGATGGGCTGAGAGGCTCTCCACCAAGTCCCTCTGGGCGCGCGCATGG 715
QY 481 AGAAGATGGACTAGGTGGACTCCAAAGTCTCTCTGAGACCTGACTCGGGGCTCTCTTT 540
Db 716 AGAAGATGGACTAGGTGGACTCCAAAGTCTCTCTGAGACCTGACTCGGGGCTCTCTTT 775
QY 541 CATGTACTCTGCCCAACGTTTTGGGGGACATCTGGGCCACAGAGGGAGGCCA-CTTGG 599
Db 776 CATGTACTCTGCCCAACGTTTTGGGGGACATCTGGGCCACAGAGGGAGGCCA-CTTGG 835
QY 600 CACCCCTCTGATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGGCC 659
Db 836 CACCCCTCTGATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGGCC 895
QY 660 AGGAGCTTTTTCAGGGCTCAGATTTGGGATGGGATGGGAGAGAGGAGAGGCGCTGGGGAGA 719
Db 896 AGGAGCTTTTTCAGGGCTCAGATTTGGGATGGGATGGGAGAGAGGAGAGGCGCTGGGGAGA 955
QY 720 AAGCGCGGACAGACAGCCCTCGCAGAGGAGCATGTGACCTCGCTACAGAGCATCTTGG 779
Db 956 AAGCGCGGACAGACAGCCCTCGCAGAGGAGCATGTGACCTCGCTACAGAGCATCTTGG 1015
QY 780 AGCAATTCCTTCAACGATGGCAGCCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA 839
Db 1016 AGCAATTCCTTCAACGATGGCAGCCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA 1075
QY 840 AGCTGAGGACATTTCCAGCAGGAGTTTCCACCCCTTCAGGAGGCGCTGGTGTTCG 899
Db 1076 AGCTGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCAGGAGGCGCTGGTGTTCG 1135
QY 900 AGCTGATCCAGTCTTACAGCGGATGCCAGGCAATGCCATGTGAGGGGCTTCCGAGTGG 959
Db 1136 AGCTGATCCAGTCTTACAGCGGATGCCAGGCAATGCCATGTGAGGGGCTTCCGAGTGG 1195
QY 960 CTATAAGCGCACGTGCTGACCATGGATGACTTGGGACCTTGTATGACAGAACTGGC 1019
Db 1196 CTATAAGCGCACGTGCTGACCATGGATGACTTGGGACCTTGTATGACAGAACTGGC 1255
QY 1020 TCAATGACCAAGTGATGAACATGTATGAGACCTGGTCATGGACACAGTCCCTGAAAAGG 1079
Db 1256 TCAATGACCAAGTGATGAACATGTATGAGACCTGGTCATGGACACAGTCCCTGAAAAGG 1315
QY 1080 TGCATTTCTTCAATAGTTTCTCTATGATAAATCCGTACCAAGGTTATGATGGGTGA 1139
Db 1316 TGCATTTCTCAATAGTTTCTCTATGATAAATCCGTACCAAGGTTATGATGGGTGA 1375
QY 1140 AAAGTGGACCAAAAACGTGGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACC 1199
Db 1376 AAAGTGGACCAAAAACGTGGACATCTTCAATAAGGAGCTACTGCTAATCCCATCCACC 1435
QY 1200 TGGAGGTGATTTGGTCCCTCATCTCTGTGTGATGTGAGGCGGAGCCACCATCATTTTGG 1259
|||||

Db 1436 TGGAGTGTCATGTGTCCTCATCTCTGTTGATGTCAGGCGAGCCACCATCATCTTTTG 1495
QY 1260 ACTCGCAGGTACCTTAAACCGCGCTGCCCTAAGCATATTCCTAAGTATCTACAGCGAG 1319
Db 1496 ACTCGCAGGTACCTTAAACCGCGCTGCCCTAAGCATATTCCTAAGTATCTACAGCGAG 1555
QY 1320 AGCGGTAAGAAAGAACCGGACTGGATTTCCACAGGCGTGGAAAGGTTACTTCAAATGA 1379
Db 1556 AGCGGTAAGAAAGAACCGGACTGGATTTCCACAGGCGTGGAAAGGTTACTTCAAATGA 1615
QY 1380 ATGTGCCAGGAGCAATATGACAGTACTGTGGTCTTTTGTGTGGAGTACTGCAAGC 1439
Db 1616 ATGTGCCAGGAGCAATATGACAGTACTGTGGTCTTTTGTGTGGAGTACTGCAAGC 1675
QY 1440 ATCTGGCCCTGCTCAGCCATTCAGCTTCACCCAGCAGACATGCCCAAACTTCGTCGCG 1499
Db 1676 ATCTGGCCCTGCTCAGCCATTCAGCTTCACCCAGCAGACATGCCCAAACTTCGTCGCG 1735
QY 1500 AGATCTACAGAGCTGTGTCACCTGCAAACTCAGTGTGTGAGCCTGCTACCCAGACCCC 1559
Db 1736 AGATCTACAGAGCTGTGTCACCTGCAAACTCAGTGTGTGAGCCTGCTACCCAGACCCC 1795
QY 1560 AAGCCCAATAATGGGAAGGAGACATGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC 1619
Db 1796 AAGCCCAATAATGGGAAGGAGACATGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC 1855
QY 1620 CTCTCTGCTCTTCCCACTCCTCCCTTTGTTTTTTCATATTTAAATGTTTCAATTTTC 1679
Db 1856 CTCTCTGCTCTTCCCACTCCTCCCTTTGTTTTTTCATATTTAAATGTTTCAATTTTC 1915
QY 1680 TGTATTTTCTTTTGTGAGAGAACTACTTGTGATTTCTGATGTGAGGGGTGGCTACA 1739
Db 1916 TGTATTTTCTTTTGTGAGAGAACTACTTGTGATTTCTGATGTGAGGGGTGGCTACA 1975
QY 1740 GAAAAGCCCTTTCTTCTCTCTTTGAGGGAGTGTGGCCCTGTGGCTGGGAGCA 1799
Db 1976 GAAAAGCCCTTTCTTCTCTCTTTGAGGGAGTGTGGCCCTGTGGCTGGGAGCA 2035
QY 1800 GTCATCTCCCTTCCCGTGCAGGAGCAGAAATCAGTGTGGGGTGGTGGCGGA 1859
Db 2036 GTCATCTCCCTTCCCGTGCAGGAGCAGAAATCAGTGTGGGGTGGTGGCGGA 2095
QY 1860 CAATAGGATCACTGCTGCCAGATCTTCAAACCTTTTATATATATATATATATATAT 1919
Db 2096 CAATAGGATCACTGCTGCCAGATCTTCAAACCTT-----TATAT 2135
QY 1920 AT 1979
Db 2136 AT 2195
QY 1980 TCAATAAAG 1988
Db 2196 TCAATAAAG 2204

RESULT 4
AX081143
LOCUS AX081143 2206 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 9 from Patent WO0109292.
ACCESSION AX081143
VERSION AX081143.1 GI:13170039
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2206)
AUTHORS Gong,L. and Yeh,E.T.
TITLE Composition and methods relating to senpi - a sentrin-specific
protease
JOURNAL Patent: WO 0109292-A 9 08-FEB-2001;
Board of Regents, The University of Texas System (US)
FEATURES Location/Qualifiers

```

source      . 1..2206
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT  491 a 612 c 605 g 498 t
ORIGIN
Query Match      92.6%; Score 1843.4; DB 6; Length 2206;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1925; Conservative 0; Mismatches 31; Indels 33; Gaps

QY  1 GCCTCCCTGTCCTCCCGACCCCTCTTTTGATGCTCTCAGCAGTGAAGAGGAGGAGGAGG 60
DB  230 GCCTCCCTGCTCCCGCGGCATCTTTTGATGCTCTCAGCTAGTGAAGAAGAGGAGGAGG 289
QY  61 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAAGTGGCAGCTTGGAGGCTGCCCCCAAGT 120
DB  290 AGGAG-----GAAATGAGGAGGAAGTAGCAGCTTGGAGGCTACCCCTTAGGT 337
QY  121 GGAGTCAGCTGGGAACCTCCACGCGCCCGCCCTTCCGCGCCCACTCATCGAAAAACCT 189
DB  338 GGGGCCAACTGGGGGCTCCACGCGCTCTCGAGCTCTCCGACCCCTCATAGAAAAACCT 397
QY  181 GCTCAGAGCCCGCCGCGAGCGCATCAGAGCCTTCGGGATGCTGCTCTACTCAAAAAGCA 240
DB  398 GCTCAGAGCCCGCGCGAGCGCATCAGAGCCTTCAGATGCTGCTCTACTCAAAAAGCA 457
QY  241 CCTCGCTGACATTCACACTGGAAGCTTTGGGGGCGCCACCGGGGCGCGCGGGGCGCTCG 300
DB  458 CCTCGCTGACATTCACACTGGAAGCTTTGGGGGCGCCACCGGGGCGCGCGGGGCGCTCG 517
QY  301 CACACCCCAAGAACCATCTTTACCCCCACAGAAGGGGGTGCAGCGCCACAGGTGCCATCCC 360
DB  518 CACACCCCAAGAACCATCTTTACCCCCACAGAAGGGGTGCAGCGCCACAGGTGCCATCCC 577
QY  361 CTGTGTTGCTTTTGACTCCCGCGGGGCGCACTCTCACCGCGGCTGGGTGCTGTAGGTG 420
DB  578 CTGTGTTGCTTTTGACTCCCGCGGGGCGCACTCTCACCGCGGCTGGGTGCTGTAGGTG 637
QY  421 CTCTCATGGCTGAGGATGGGTGAGAGGCTCTCACCACTGCCCTCTGGCGCCCGCCATGG 480
DB  638 CTCTCATGGCTGAGGATGGGTGAGAGGCTCTCACCACTGCCCTCTGGCGCCCGCCATGG 697
QY  481 AGGAAGATGGACTCAGGTGGACTCCAAGTCTCCTCTGGACCCCTGACTCGGGGCTCCTTT 540
DB  698 AGGAAGATGGACTCAGGTGGACTCCAAGTCTCCTCTGGACCCCTGACTCGGGGCTCCTTT 757
QY  541 CATGTACTCTGCCCAACGGTTTGGGGGACAATCTGGGGCGAGAGGGGAGGCGCA-CTTGG 599
DB  758 CATGTACTCTGCCCAACGGTTTGGGGGACAATCTGGGGCGAGAGGGGAGGCGCACTTGG 817
QY  600 CACCCCTGATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGACCACTATGGGCC 659
DB  818 CACCCCTGATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGACCACTATGGGCC 877
QY  660 AGGAGCTTTTTCAGGCTCAGATTTGGGCATGGCAGAGAGGCGAGAGGCGCTGGGAGA 719
DB  878 AGGAGCTTTTTCAGGCTCAGATTTGGGCATGGCAGAGAGGCGAGAGGCGCTGGGAGA 937
QY  720 AAGCGGCGCAGCAGACGCCCTCTGGAGAGGAGCATGTGACCTCGCTACAGAGCATCTTGG 779
DB  938 AAGCGGCGCAGCAGACGCCCTCTGGAGAGGAGCATGTGACCTCGCTACAGAGCATCTTGG 997
QY  780 ACGAATTCCTTCAAGCGTATGCGAGCCTCATACCCCTCAGCAGCTGATCAGCTACTAGAGA 839
DB  998 ACGAATTCCTTCAAGCGTATGCGAGCCTCATACCCCTCAGCAGCTGATCAGCTACTAGAGA 1057
QY  840 AGCTGGAGGACATTTTCCAGCAGGAGTTTTCACCCCTTCCAGGAAGGGCCTGGTGTTCG 899
DB  1058 AGCTGGAGGACATTTTCCAGCAGGAGTTTTCACCCCTTCCAGGAAGGGCCTGGTGTTCG 1117
QY  900 AGCTGATCAGTCTTACCAAGCGGATGCCAGGCAATGCCATGGTGAGGGGCTTCCAGTGG 959
DB  1118 AGCTGATCAGTCTTACCAAGCGGATGCCAGGCAATGCCATGGTGAGGGGCTTCCAGTGG 1177

```

RESULT 5	HSMB00198	1782 bp	mRNA	linear	PRI 10-MAR-2001
LOCUS	Homo sapiens mRNA; cDNA DKF2p586K0919 (from clone DKF2p586K0919); complete cds.				
DEFINITION	AL050283				
ACCESSION	AL050283.1	GI:4886466			
VERSION					
KEYWORDS	human.				
SOURCE					
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 1782) Wiemann, S., Weil, B., Wellenreuther, R., Gassenhuber, J., Glassl, S., Ansong, W., Boecker, M., Bloeker, H., Bauersachs, S., Blum, H., Lauber, J., Duesterhoeft, A., Beyer, A., Koehler, K., Strack, N., Mewes, H. W., Othenwaelder, B., Obermaier, B., Tampe, J., Heubner, D., Wambutt, R., Korn, B., Klein, M. and Poustka, A.				
TITLE	Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs				
JOURNAL	Genome Res. 11 (3), 422-435 (2001)				
PUBMED	11230166				
REFERENCE	2 (bases 1 to 1782) Wambutt, R., Heubner, D., Mewes, H. W., Gassenhuber, J. and Wiemann, S.				
AUTHORS	Direct Submission				
TITLE	Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY				
JOURNAL	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.				
COMMENT	This clone (DKF2p586K0919) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cbna/. Location/Qualifiers 1. .1782 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="DKF2p586K0919" /tissue_type="uterus" /clone_lib="586 (synonym: hutel). Vector pSport1; host DH10B; sites NotI + SalI/MluI" /dev_stage="adult" 204. .1319 /gene="DKF2p586K0919" 204. .1319 /gene="DKF2p586K0919" /note="similarity to S.pombe SPBC19G7.09" /codon_start=1 /product="hypothetical protein" /protein_id="CAB43384.1" /db_xref="GI:4886467" /db_xref="SPTREMBL:Q9Y3W9" /translation="NAEDGVRSPPVPPMEEDGLRWTPKSPDLPDPSGLLSCTLPN GFGGSGPGESESLAPPDASILLISNVCSIGDHVAFLEFGSDLSNAEAREPGEKAGQ HSPLEBHWTCQVSLDEFLTQYSGLSIFLSTDEVVEKLEDFQGEFSPSRGLVQL IQSYRPMGNAMVRGFRVAYKRHLVTMDDLGLTYGQNLNDQVMNMGDLYNDVTPEK VHFFSFFYDKLRKGYGVKRWTKNVDIFNKLLELLIPHLEVHWSLISVDYRRRTIT YFDSQRTNRPCKPHIAKYLAQEAQVKDRDLDFHOGKGYFKMNVARQNNDSDGAFVL QYCKHLALSQPSFTQDDMPKLRQIYKELCHCKLTV" polysite 1741. .1746 polysite 1765 BASE COUNT 419 a 466 c 475 g 422 t ORIGIN				
Query Match	85.0%; Score 1692.4; DB 9; Length 1782;				
Best Local Similarity	98.7%; Pred. No. 0;				
Matches 1734; Conservative	0; Mismatches 1; Indels 21; Gaps 2;				
QY 234	AAAAGCACCTCGTGACATTCACACTGGAAGCTTTGGGGCGCGCACCGGGCGCGGCGG 293				

Db	12	AAAAGCACCTCGTGACATTCACACTGGAAGCTTTGGGGCGCGCACCGGGCGCGGCGG 71
QY	294	GGCCTCGCACACCCCAAGAACCATCTTTTACCCCAAGAGGGGTGCGAGCCACAGGTG 353
Db	72	GGCCTCGCACACCCCAAGAACCATCTTTTACCCCAAGAGGGGTGCGAGCCACAGGTG 131
QY	354	CCATCCCCCTGTGTGCTTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGCTGGGTCTG 413
Db	132	CCATCCCCCTGTGTGCTTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGCTGGGTCTG 191
QY	414	CTAGTGTCTCTCATGTGCTGAGGATGGGTGAGAGGTCTCTCCACAGTGCCTCTTGGGCC 473
Db	192	CTAGTGTCTCTCATGTGCTGAGGATGGGTGAGAGGTCTCTCCACAGTGCCTCTTGGGCC 251
QY	474	CCCATGGAGGAAGATGCATCAGGTGGACTCCAAAGTCTCTCTGACCCCTGACTCGGGC 533
Db	252	CCCATGGAGGAAGATGCATCAGGTGGACTCCAAAGTCTCTCTGACCCCTGACTCGGGC 311
QY	534	CTCCTTTTCATGTACTCTGCCCAACGGTTTGGGGGACAATCTGGGCCAGAAAGGGAGCGC 593
Db	312	CTCCTTTTCATGTACTCTGCCCAACGGTTTGGGGGACAATCTGGGCCAGAAAGGGAGCGC 371
QY	594	A-CTTGGCACCCCTGATGCCAGCATCTCATACAGCAATGTGTGACGATCGGGGCCAT 652
Db	372	AGCTTGGCACCCCTGATGCCAGCATCTCATACAGCAATGTGTGACGATCGGGGCCAT 431
QY	653	GTGGCCCCAGAGCTTTTTCAGGGCTCAGATTGGGGATGGCAGAGAGAGAGGCCT 712
Db	432	GTGGCCCCAGAGCTTTTTCAGGGCTCAGATTGGGGATGGCAGAGAGAGAGGCCT 491
QY	713	GGGGAAAAGCGCGCCAGCACACGCCCTCGCAGAGAGAGCATGTGACCTGCGTACAGAGC 772
Db	492	GGGGAAAAGCGCGCCAGCACACGCCCTCGCAGAGAGAGCATGTGACCTGCGTACAGAGC 551
QY	773	ATCTTGGACGAATTCCTTCAAAGTATGGCAGCCTCATACCCCTCAGCACTGTATGAGGTA 832
Db	552	ATCTTGGACGAATTCCTTCAAAGTATGGCAGCCTCATACCCCTCAGCACTGTATGAGGTA 611
QY	833	GTAGAGAAGCTGGAGGACATTTTCCAGCAGAGAGTTTCCACCCCTTCCAGGAAGGCCCTG 892
Db	612	GTAGAGAAGCTGGAGGACATTTTCCAGCAGAGAGTTTCCACCCCTTCCAGGAAGGCCCTG 671
QY	893	GTGTTTCAGCTGATCCAGCTTTTACCAGCGGATGCCAGGAATGCCATGTGTGAGGGCTTC 952
Db	672	GTGTTTCAGCTGATCCAGCTTTTACCAGCGGATGCCAGGAATGCCATGTGTGAGGGCTTC 731
QY	953	CGAGTGGCTTATAAGCGGCACGCTGTGACCATGATGACTTGGGGACCTTGTATGGACAG 1012
Db	732	CGAGTGGCTTATAAGCGGCACGCTGTGACCATGATGACTTGGGGACCTTGTATGGACAG 791
QY	1013	AACTGGCTCAATGACCCAGGTGATGAACATGTATGGAGACCTGGTCATGACACAGTCCCT 1072
Db	792	AACTGGCTCAATGACCCAGGTGATGAACATGTATGGAGACCTGGTCATGACACAGTCCCT 851
QY	1073	GAAGAAGTGCATTTCTTCAATAGTTTCTTATGATAAACTCCGTACCAGGGCTTATGAT 1132
Db	852	GAAGAAGTGCATTTCTTCAATAGTTTCTTATGATAAACTCCGTACCAGGGCTTATGAT 911
QY	1133	GGGGTGAAGAAGTGGACCAAAAACGTGGACATCTTCAATAAGAGAGTACTGTCTAAATCCC 1192
Db	912	GGGGTGAAGAAGTGGACCAAAAACGTGGACATCTTCAATAAGAGAGTACTGTCTAAATCCC 971
QY	1193	ATCCACCTGGAGTGCATTTGGTGGCTTCATCTCTGTGTGATGTGAGGCGACGCCACCATCAC 1252
Db	972	ATCCACCTGGAGTGCATTTGGTGGCTTCATCTCTGTGTGATGTGAGGCGACGCCACCATCAC 1031
QY	1253	TATTTTGACTCGCAGGTACCCCTTAAACCGCGCTCCCTTAAAGCATATTGCCAAGTATCTA 1312
Db	1032	TATTTTGACTCGCAGGTACCCCTTAAACCGCGCTCCCTTAAAGCATATTGCCAAGTATCTA 1091
QY	1313	CAGGCAGAGCGGTAAAGAAAGAACCCGACTTGGATTTCACAGCGGCTGGAAGAGTTACTTC 1372

Db	1092	CAGCAGAGCGGTAAGAAAGACCGACTGGATTTCACCAGGCGTGGAAAGGTTACTTC	1151
Qy	1373	AAATGAATGTGGCAGGAGCAATAATGACAGTCACTGTGTGCTTTGTGTTGCAGTAC	1432
Db	1152	AAATGAATGTGGCAGGAGCAATAATGACAGTCACTGTGTGCTTTGTGTTGCAGTAC	1211
Qy	1433	TGCAAGCATCTGGCCCTGTCTCAGCCATTACGCCCTTCCACCAGCAGACATGCCCAAAC	1492
Db	1212	TGCAAGCATCTGGCCCTGTCTCAGCCATTACGCCCTTCCACCAGCAGACATGCCCAAAC	1271
Qy	1493	CGTCGGAGATCTACAGGAGCTGTGTCAGTCACTGCAAACTCACTGTGTGAGCCTCGTAC	1552
Db	1272	CGTCGGAGATCTACAGGAGCTGTGTCAGTCACTGCAAACTCACTGTGTGAGCCTCGTAC	1331
Qy	1553	AGACCCCAAGCCCAATAATGGAGGAGAGACATGGGAGTCCCTTCCCAAGAAACTCCAG	1612
Db	1332	AGACCCCAAGCCCAATAATGGAGGAGAGACATGGGAGTCCCTTCCCAAGAAACTCCAG	1391
Qy	1613	TCCTTTCTCTTTCCTCTTCCCACTCACTTCCCTTTGGTTTTTTCATATTTAAATGTTT	1672
Db	1392	TCCTTTCTCTTTCCTCTTCCCACTCACTTCCCTTTGGTTTTTTCATATTTAAATGTTT	1451
Qy	1673	CAATTTCTGTATTTTTTTTCTTTGAGAGATACTTGTGATTTCTGATGTGCAGGGGT	1732
Db	1452	CAATTTCTGTATTTTTTTTCTTTGAGAGATACTTGTGATTTCTGATGTGCAGGGGT	1511
Qy	1733	GGCTACAGAAAGCCCTTTCTCTCTGTTTGCAGGGGAGTGTGGCCCTGTGGCCTGGG	1792
Db	1512	GGCTACAGAAAGCCCTTTCTCTCTGTTTGCAGGGGAGTGTGGCCCTGTGGCCTGGG	1571
Qy	1793	TGGAGCAGTCATCTCCCTTCCCTGTCAGGAGCAGGAAATCACTGTGTGGGGTGGT	1852
Db	1572	TGGAGCAGTCATCTCCCTTCCCTGTCAGGAGCAGGAAATCACTGTGTGGGGTGGT	1631
Qy	1853	GGCGGACATAGATCACTGCGCTGCCAGATCTTCAAACTTTTATATATATATATATA	1912
Db	1632	GGCGGACATAGATCACTGCGCTGCCAGATCTTCAAACTTTTATATATATATATATA	1673
Qy	1913	TAT	1972
Db	1674	--TAT	1731
Qy	1732	GCTCTGTCATAAAG 1988	
Db	1732	GCTCTGTCATAAAG 1747	
RESULT 6			
LOCUS	AY008764	2174 bp	linear ROD 21-NOV-2000
DEFINITION	Mus musculus sentrin/SUMO-specific protease (SEN3) mRNA, complete cds.		
ACCESSION	AY008764		
VERSION	AY008764.1	GI:11245812	
KEYWORDS	Mus musculus.		
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Yeh, E.T., Gong, L. and Kamitani, T.		
TITLE	Ubiquitin-like proteins: new wines in new bottles		
JOURNAL	Gene 248 (1-2), 1-14 (2000)		
MEDLINE	20267842		
PUBMED	10806345		
REFERENCE	2 (bases 1 to 2174)		
AUTHORS	Gong, L. and Yeh, E.T.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-OCT-2000) Institute of Molecular Medicine, University of Texas-Houston, 2121 W. Holcombe Blvd. 928, Houston, TX 77030, USA		
FEATURES	Location/Qualifiers		
source	1. .2174		

gene	/organism="Mus musculus"			
	/strain="C57BL/6J"			
	/db_xref="taxon:10090"			
	/chromosome="11"			
	1. .2174			
CDS	/gene="SEN3"			
	52. .1758			
	/gene="SEN3"			
	/codon_start=1			
	/product="sentrin/SUMO-specific protease"			
	/protein_id="AAG33253.1"			
	/db_xref="GI:11245813"			
	/translation="MKETIGTSGWGPPEPPGTTYNPRRLRLPPLPKPKRLKSGG			
	GPGDPGSGTTPTRLPAPRPSFDASAESEEEDEEEVAANRLPPRWQLGAS			
	QSRALRPSHRKTCORRRAMRFAOMLLYSKSTSLTFHWKLMGRGRRLNAHPKN			
	HUSPQGGATPQVPSCCRPDPRGLPPPLGLLALMAEDMRGSPFPVSGPPEED			
	GLRTPKSLDPSGLSCTLPNGFGLSGEGERSLAPDASILISNVCISGDHVAQ			
	ELFQSDLGIAEADRTKAGQHSRLREHVTQVSIIDFLQTYGSLTIPLDLVY			
	EKLEDIFQGEFTSPSRKSLVQLIQSRMPGNAMVGRFVSKRVKLTMDIDLTGLY			
	ONWLNQVMNMYGDLVMDTVPEKVHFENSFYDKLRTKGYDGVKRWTKNVDIENKLL			
	LPIHLHVHSLISVDVRRRTITYFDORTLAPRCPKIAKYLOAEAVKDRLDHFQG			
	WKGYFKMNVARQNNSDCGAFVLQYCKHLALSQPFSTQDDMPKLRQIKYELCHCKL			
	TV"			
BASE COUNT	513 a	585 c	570 g	506 t
ORIGIN				
	Query Match 73.3%; Score 1459.2; DB 10; Length 2174;			
	Best Local Similarity 87.5%; Pred. No. 0;			
	Matches 1684; Conservative 0; Mismatches 213; Indels 27; Gaps 7;			
Qy	1	GCCTCCCTGTCGCCCGACCCCTCTTTTGATGCCTCAGCAAGTGAAGAGAGGAGGAAGAGG	60	
Db	230	GCCTCCCTGTCGCCCGACCCCTCTTTTGATGCCTCAGTGAAGAGAGGAGGAAG	289	
Qy	61	AGGAGGAGGAGGATGAAGATGAAGAGAGGAGGAGTGGCAGCTTGGAGGCTGCCCCAGAT	120	
Db	290	AGGAG-----GAAGATGAGGAGGAGTAGCAGCTTGGAGGCTACCCCTAGT	337	
Qy	121	GGAGTCAGCTGGGAACCTCCACAGGGCCCGCCCTTCCCGCCACATCATCGAAAAACCT	180	
Db	338	GGGGCCAACTGGGGGGCTCCACGGCTCTCGAGCTCTCCGACCCCTCTCATAGAAAAACCT	397	
Qy	181	GCTCACAGCGCCCGCGGAGCCATGAGAGCCTTCCGAGTCTCTCTACTCAAAAAGCA	240	
Db	398	GCTCACAGCGCCCGCGGAGCCATGAGAGCCTTCCAGATGCTCTACTCAAAAAGCA	457	
Qy	241	CCTCGCTGACATTCCTACTGGAAGCTTTGGGGGGCCACCGGGGGCGGGGGGCGCTCG	300	
Db	458	CCTCGCTGACATTCCTACTGGAAGCTTTGGGGGGCCACCGGGGGCGGGGGGCGCTCG	517	
Qy	301	CACACCCCAAGAACCATCTTTCACCCAGCAAGGGGTGCGAGCCACAGTGCCTATCCC	360	
Db	518	CGCACCCCAAGAACCATCTCTCACCCAGCAAGGGGTGCAACCCACAGTGCCTATCAC	577	
Qy	361	CCTGTTGCTGCTTTTGTAGTCCCGGGGGCCACTTCCACCCCGGGCTGGGTCTGTAGGTG	420	
Db	578	CCTGCTGCTGTTTGTAGTCCCGGGGGCTACCCCGGGCTGGGTCTGTAGGTG	637	
Qy	421	CTCTCATGCTGAGGATGGGGTGAAGGGTCTCCACAGTGCCTCTGGGGCCCCCATGG	480	
Db	638	CTCTCATGCGAGAGATGGGATGAGAGGGTCTCCACAGTGCCTCTGGGGCCCCCATGG	697	
Qy	481	AGGAAGATGGACTCAGTGGACTCCAAAGTCTCTCTGGACCTGACTCGGGGCTCCTTT	540	
Db	698	AGGAAGATGGACTAAGTGGAGTGGACCCCAAGTCTCTCTGGACCTGACTCGGGGCTCCTCT	757	
Qy	541	CATGTACTCTGCCCAACGGTTTTTGGGGGCAATCTTGGGCCAGAGGGAGCGCA-CTTGG	599	
Db	758	CATGTACTCTACCAATGGTTTTTGGGGGACTGTCTGTCTGAAGGGAGCGCAGCTGG	817	
Qy	600	CACCCCTGATGCCAGATCCTCATCAGCAANTGTGTGACAGCATCGGGGACCATGTGGCCC	659	
Db	818	CACCCCTGATGCCAGCATCTCATCAGTAATGTGTGACATTTGAGAGACAGCTGGGCTC	877	

QY	660	AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAAAGAGCGCAGAGAGCCCTGGGGAGA	719
Db	878		
		AGGAACCTTTTTCAGAGCTCTGACTTTGGGCATTTGCAGAAGAGCGCAGATCGGACTCGGGAGA	937
QY	720		
		AAGCGGGCAGCACAGCCCCCTTCGCAGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGG	779
Db	938		
		AAGCTGGCCAGCATAGCCCCCTTCGGGAGAGAACATGTGACCTGTGTGCAGATGATCTTAG	997
QY	780		
		AGAAATTCCTTCAAAACGATATGGCAGCCCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA	839
Db	998		
		ATGAATTCCTTCAAAACTTATGGCAGCCCTCATCCCTCTCAGCACTGATGAAGTTGTAGAGA	1057
QY	840		
		AGCTGGAGACATTTTTCAGCAGGAGTTTTCCACCCCTTCAGAAAGGGCCTGGTGTTC	899
Db	1058		
		AGTTGGAGACATTTTTCAGCAGGAGTTCCTACACCCCTCAAGGAAGAGTCTGGTACTAC	1117
QY	900		
		AGCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGGTCAGGGGCTTCGAGATGG	959
Db	1118		
		AGCTGATCCAGTCTTATCAGCGGATGCCAGCAACGCTATGGTAAGGGGCTTCGGGTAT	1177
QY	960		
		CTTATAAGCGGCACGTGCTGACCATGGATGACTTTGGGACCTTGTATGGACAGAACTGGC	1019
Db	1178		
		CTTATAAGCGACACCTGGCTCACATGGATGACTTTGGGTACCTTATATGGACAGAACTGGC	1237
QY	1020		
		TCAATGACCAGGTGATGAACATGATGAGACCTGGTCATGGACACAGTCCCTGAAAAGG	1079
Db	1238		
		TCAATGACCAGGTGATGAACATGATGAGAGTCTGGTCATGGACACAGTCCCTGAAAAGG	1297
QY	1080		
		TGCATTTCTCAATAGTTCCTTCTATGATAAACTCCGTACCAGGGTATGATGGCGTGA	1139
Db	1298		
		TGCATTTCTCAACAGCTCTCTCATGATAAATCCGTACCAAGGTTATGATGGGGTAA	1357
QY	1140		
		AAAGTGGACAAAACGTGGACATCTTCAATAAGAGGCTACTGCTAATCCCCATCCACC	1199
Db	1358		
		AGAGTGGACAAAATGTGGACATCTTCAATAAGAGGAATTAAGTCTAATCCCCATCCATC	1417
QY	1200		
		TGGAGTGCATTGGTCCCTCATCTCTGTGTGATGTGAGCGAGCACCATCACCTATTTTG	1259
Db	1418		
		TGGAGTGCATCTGGTCCCTTATCTCAGTGTGATGTAAGCGAGCTACCATCACCTATTTTG	1477
QY	1260		
		ACTCGACGGTACCCTAAACCGCCCTGCCCTTAAGCATATTGCCAAGTATCTACAGGCAG	1319
Db	1478		
		ACTCCAGCGAACTCTAAATCGCCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAG	1537
QY	1320		
		AGCGGTAAAGAAAGACCGACTGGATTTCCACAGGGGCTGAAAGGTTACTTCAAAATGA	1379
Db	1538		
		AGCGAGTCAAAAAGACCGGACTTGGACTTCCATCAGGGCTGGAAGGTTACTTCAAAATGA	1597
QY	1380		
		ATGTGCCAGGCAGAAATATGACAGTGAATGTGGTGTCTTGTCTGCAGTACTGCAAGC	1439
Db	1598		
		ATGTGCCAGGCAGAAATATGACAGTGAATGTGGTGTCTTGTATACAGTACTGCAAGC	1657
QY	1440		
		ATCTGGCCCTGTCTCAGGCATTCAGCTTTCACCCAGCAGGACATGCCCAAACCTTCGTCCGC	1499
Db	1658		
		ACTGGCCCTGTCTCAGGCATTCAGCTTTCACCCAGCAGGACATGCCCAAACCTTCGTCCGC	1717
QY	1500		
		AGATCTCAAGAGAGCTGTCTCAGTCAAACTCATCTGTGTGAGCCTCGTACCCAGACCCC	1559
Db	1718		
		AGATCTCAAGGAAGTGTGTCTACTGCAAACTCACTGTGTGAGCCTCATACCCAGACCCCT	1777
QY	1560		
		AAGCCCATAAATGGGAAGGGAGACATGGAGTCCCTTCCCAAGAACTCCAGTTCCTTC	1619
Db	1778		
		TAGCCCATCAGTGGCAGG--GGAAGGGAGACCTTCCCAAGAACTCTAGTTTCCT---	1832
QY	1620		
		CTCTCTTGCCCTCTTCCCATCTCCTTGGTTTTTCATATTAAATGTTTTCAATTTC	1679
Db	1833		
		-----TTTTTGTCTCTACCCAAATCCCTTTGGATTTTCATATTAAATGTTT-GGTTTC	1886
QY	1680		
		TGTAATTTTTTTTTCTTTGAGAGAAATACTGTTGATTTCTGATGTGACGGGGTGGCTACA	1739
Db	1887		
		TGTAATTTTTTTTTTGTCTGCGAGTATATGATTTCTGATGGAAGGAATGGCCACA	1946

Qy	1740	GAAAAGCCCTTTCCTCCTTGTTGACAGGGAGTGTGGCCCTGTGCCTCGGTGGAGCA	1799
Dd	1947	TAAAGAACAAGCCTCTTTCTCCCTCATATAGAGGAATGTTCCTTGTGCCTTAGTGGAGCA	2006
Qy	1800	GTCACTCTCCCCTTCCCCTGTGACGAGGACAGAAAAATCAGTGTGGGGTGTGGCGGA	1859
Dd	2007	GCCAT--TCCCTCCTTCATGTTTCAGAGGCCAGAAAACAGTG-TGGGGGTGCTGCAGCGC	2063
Qy	1860	CAATPAGGACTACTGCCCTGCCAATCTTCAAACCTTTTTATATATATATATATATATATAT	1919
Dd	2064	CAAAGGATTACTGCCCTGCCGATCTTCAAACCTTTTTATATATATAATAATATATAA	2123
Qy	1920	ATAT 1923	
Dd	2124	AAAT 2127	
RESULT 7			
AF194031	AF194031	1707 bp	mRNA linear ROD 01-NOV-2000
LOCUS	Mus musculus SMT3 isopeptidase 1 (Smt3ip)		mRNA, complete cds.
DEFINITION	AF194031		
ACCESSION	AF194031.1	GI:11066007	
VERSION			
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE			
AUTHORS	Nishida,T., Tanaka,H. and Yasuda,H.		
TITLE	A novel mammalian Smt3-specific isopeptidase 1 (SMT3IP1) localized in the nucleolus at interphase		
JOURNAL	Eur. J. Biochem. 267 (21), 6423-6427 (2000)		
MEDLINE	20485517		
PUBMED	11029585		
REFERENCE			
AUTHORS	Nishida,T., Tanaka,H. and Yasuda,H.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-OCT-1999) School of Life Science, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan		
FEATURES			
source	Location/Qualifiers		
	1..1707		
	/organism="Mus musculus"		
gene	/db_xref="taxon:10090"		
	1..1707		
	/gene="Smt3ip"		
CDS	1..1707		
	/codon_start=1		
	/product="SMT3 isopeptidase 1"		
	/protein_id="AAC28418.1"		
	/db_xref="GI:11066008"		
	/translation="MKETIQGTSGWPEPPPGTTSNPRRRLRWLPPLPKPLKSGGFGFDPSGVTVTRLPAPRPFDASASEEEEEEEVEEVAAWRLPPRWGLQGSORSLARSHRTCSQRRRRAMAFOMLYSKTSLTFHWKLGRHRRRNLAHPKNHLSPQEGGATFPQVPSCCRDPSRGLPPRLGLLALMAEDMGSRPVPPSGPMEDHLSPGKSLPDPSGLLSCTLPNGFGLSGPEGERSIAPDASITILSNVCISGDHVQAELFQSDSIQTAEARDTEGARGEKSPUREEHVTVCVSILDEFLQTYGSLIPLSTDEVYIKNLIDFQQEFTSPRSKSIVLIQLIQSYRPNMGNAMRGFRVYSKKRHVLTWDLDLTLGKQNWLDQVMNYGDLMVDPERKHFFNSFYDKLRKTGYDGVRKWTKNVDIIFNKELLIPITLHEVHWSLISVDVRRRTITYFDSQPTLNRRCPKHIAKYLAQEAAYKKDRIDFHOGWKPYEFKMVNVARQNDSDCGAFVLQYCKHIALSQPFSTQDMPKLRRIYKELCHCKCLTV"		
BASE COUNT	406 a	477 c	461 g 363 t
ORIGIN			
Query Match	64.2%	Score 1279;	DB 10; Length 1707;
Best Local Similarity	90.4%	Pred. No. 1.4e-303;	
Matches 139;	Conservative	0; Mismatches 135;	Indels 13; Gaps 2;
Qy	1	GCCTCCCTGTCCCCGACCCTCTTTTGTATGCTCAGCAAGTGAAGAGGAGGAGG	60

Db 179 GCCTCCTGCCCGCCCATCTTTTATGCCCTCAGCTAGTGAAGAGAGAGAGAG 238
QY 61 AGGAGGAGGAGGATGAAGATGAAGAGAGAGAGTGGCAGCTTGGAGCTCCCGCCCAAGAT 120
Db 239 AGAG-----GAAGATGAGAGAGAGTAGCAGCTTGGAGCTACCCCTTAGGT 286
QY 121 GGAGTCAGCTGGGAACCTCCAGCGGCCCGCCCTTCCCGCCCACTCATFCGAAAAACCT 180
Db 287 GGGGCCAACTGGGGGCTCCAGCGCTCTCGAGCTCTCCGACCTCTCATAGAAAAACCT 346
QY 181 GCTCACAGCGCCCGCCGAGGACCATGAGACCTTCCGGATGCTGCTACTCAAAAGCA 240
Db 347 GCTCACAGCGCCCGCCGAGGACCATGAGACCTTCCAGATGCTGCTACTCAAAAGCA 406
QY 241 CCTCGCTGACATTCACCTGGAAGCTTTGGGGGGCGCCACCGGGCGCGCGGGGCTCG 300
Db 407 CTTGCTGACATTCACCTGGAAGCTTTGGGGGGCGCCACCGAGCGCGGGCGGAACCTCG 466
QY 301 CACACCCCAAGAACCATCTTTACCCCAAGCAAGGGGGTGGAGCCACAGGTGCCATCC 360
Db 467 CGCACCCCAAGAACCATCTCTCACCCCAAGGAAGGGGTGCAACGCCACAGGTGCCATCAC 526
QY 361 CTTGCTGCTCTTTGACTCCCGCGGGGCCACTTCCACCCCGCTGGTCTCTAGGTG 420
Db 527 CTTGCTGCTCTTTGACTCCCGCGGGGCTACCCCGGCTGGTCTCTAGGTG 586
QY 421 CTCTCATGGCTGAGGATGGGTGAGAGGGTCTCCACAGTGCCCTCTGGGCCCGCCATGG 480
Db 587 CTCTCATGGCAGAGGATGGATGAGAGGTCTCCACAGTGCCCTCTGGGCCCGCCATGG 646
QY 481 AGGAAGTGCATCAGGTGGACTCCAAAGTCTCTCTGGACCTGACTCGGGGCTCTCTT 540
Db 647 AGGAAGTGCATCAGGTGGACCCCAAGTCTCTCTGGACCTGACTCGGGGCTCTCTT 706
QY 541 CATGTACTCTGCCCAACGGTTTTGGGGGCAATCTCTGGGCCAGAGGGAGCGCA-CTTGG 599
Db 707 CATGTACTCTFACCAATGTTTTGGGGGACTGCTGTGCTGAAGGGAGCGCAGGCTGG 766
QY 600 CACCCCTGATGCGCAGCATCCTCATCAGCAATGTGTCAGCATCGGGGACCATGTGGCCC 659
Db 767 CACCCCTGATGCGCAGCATCCTCATCAGTAATGTGTGTCAGCATTTGGAGCCAGTGGCTC 826
QY 660 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGCGCAAGAGGAGGAGCGCTTGGGGAGA 719
Db 827 AGGAACCTTTTTCAGAGCTCTGACTTGGGCATTCGAGAAGAGGAGCATCGGACTGGGGAGA 886
QY 720 AAGCGGCCAGCACGCCCTCGGAGGAGCATGTGACCTCGGTACAGAGCATCTTG 779
Db 887 AAGTGGCCAGCATAGCCCTTTCGGGAGGAACATGTGACCTGTGTGAGAGTATCTTAG 946
QY 780 ACGAATTCCTTCAACGATGCGAGGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA 839
Db 947 ATGAATTCCTTCAACGATGCGAGGCTCATACCCCTCAGCACTGATGAGGTAGTAGAGA 1006
QY 840 AGCTGGAGGACATTTCCAGCAGAGTTTTCACCCCTTCCAGGAAGGGGCTGGTGTTC 899
Db 1007 AGTTGGAGGACATTTCCAGCAGAGTTTCTCTACACCCCTCAAGGAAGAGTCTGGTACTAC 1066
QY 900 AGCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGGTCAGGGGCTCCGAGTGG 959
Db 1067 AGCTGATCCAGTCTTATCAGCGGATCCAGGCAACCTATGGTAAGGGGCTCCGGGTAT 1126
QY 960 CTTATAAGCGGCGACGTGCTGACCATGGATGACTTTGGGACCTTGTATGGACAACTGGC 1019
Db 1127 CTTATAGCGACACGTGCTCACCATGGATGACTTGGGTACCTTATATGGACAACTGGC 1186
QY 1020 TCAATGACCAAGGTGATGAACATGATGGAGCTTGTTCATGGACACAGCTCCCTGAAAAGG 1079
Db 1187 TCAATGACCAAGGTGATGAACATGATGGAGATCTGTCATGGACACAGCTCCCTGAAAAGG 1246
QY 1080 TGCATTCTTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTGA 1139
Db 1247 TGCATTCTTCAACAGCTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGGTAA 1306

QY 1140 AAAGGTGGACCAAAAACGTGGACATCTTCAATAAGGAGCTACTGCTTAATCCCATCCACC 1199
Db 1307 AGAGGTGGACCAAAAATGTGGACATCTTCAATAAGGAATTAAGTCTTAATCCCATCCATC 1366
QY 1200 TGGAGTGCATTTGGTCCCTCATCTCTGTGTGATGTAGGCGAGCAGCACCATTATTTTG 1259
Db 1367 TGGAGTGCATTTGGTCCCTTATCTCAGTTGTATGTAAGGGAGCTTACCATCACCATTATTTG 1426
QY 1260 ACTCGCAGCGTACCCCTAAACCGCGCTGCCCTTAAGCATATTGCCAAGTACTACAGCGAG 1319
Db 1427 ACTCCAGCGAACTTAAATCGCGCTGCCCTTAAGCATATTGCCAAGTACTACAGCGAG 1486
QY 1320 AGCGGTAAAGAAAGACCGACTGGATTTCACCGAGCGCTGGAAGGTTACTTCAAAATGA 1379
Db 1487 AGGCACTCAAAAAGAACCGACTGGACTTCCATCAGGCTGGAAGGTTACTTCAAAATGA 1546
QY 1380 ATGTGCGCCAGGAGCAATAATGACAGTACTGTGGTCTCTTTGTGTGTCAGTACTGCAAGC 1439
Db 1547 ATGTGCGCCAGGAGCAATAATGACAGTACTGTGGTCTCTTTGTATTTACAGTACTGCAAGC 1606
QY 1440 ATCTGCGCTGTCTCAGCCATTACAGTTCACCGCAGCAGCATGCCCAAACTTCGTCGCG 1499
Db 1607 ACCTGCGCTGTCTCAGCCATTACAGTTCACCGCAGGAGCATGCCCAAACTCCGTCGTC 1666
QY 1500 AGATCTACAAAGGAGCTGTGTCTACTGCAAACTCACTGTGCA 1540
Db 1667 AGATCTACAAAGGAGCTGTGTCTACTGCAAACTCACTGTGCA 1707

RESULT 8

AC016876/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

177703 bp DNA linear HTG 06-AUG-2002
Homo sapiens clone RP11-186B7, *** SEQUENCING IN PROGRESS ***, 10
unordered pieces.

AC016876
AC016876.5 GI:21313830
HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.

Homo sapiens
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 177703)

Birren,B., Nusbaum,C. and Lander,E.

Homo sapiens chromosome, clone RP11-186B7

Unpublished

2 (bases 1 to 177703)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Kleh,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 177703)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Cammarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 159951 bases at least Q40
Consensus quality: 161126 bases at least Q30
Consensus quality: 161900 bases at least Q20
Insert size: 185000; agarose-fp
Quality coverage: 7.39x in Q20 bases; sum-of-contigs
Quality coverage: 8.34x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3137: contig of 3137 bp in length
* 3138 3237: gap of unknown length
* 3238 13872: contig of 10635 bp in length
* 13873 13972: gap of unknown length
* 13973 19270: contig of 5298 bp in length
* 19271 19370: gap of unknown length
* 19371 30584: contig of 11214 bp in length
* 30585 30684: gap of unknown length
* 30685 45302: contig of 14618 bp in length
* 45303 45402: gap of unknown length
* 45403 76670: contig of 31268 bp in length
* 76671 76770: gap of unknown length
* 76771 164504: contig of 87734 bp in length.
FEATURES
source
1. .164504
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-149M23"
/clone_lib="RP43"
1. .3137
/note="assembly_fragment"
3238. .13872
/note="assembly_fragment"
13973. .19270
/note="assembly_fragment"
19371. .30584
/note="assembly_fragment"
clone_end:SP6
vector_side:right
30685. .45302
/note="assembly_fragment"
45403. .76670
/note="assembly_fragment"
76771. .164504
/note="assembly_fragment"
BASE COUNT 38842 a 42831 c 41755 g 40472 t 604 others
ORIGIN
Query Match 26.6%; Score 530.4; DB 2; Length 164504;
Best Local Similarity 99.8%; Pred. No. 2.2e-119;
Matches 531; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCCTCCCTGTCGCCCGACCCCTCTTTTGATGCCCTCAGCAGTGAAGAGGAGGAAGAGG 60
Db 25116 GCCTCCCTGTCGCCCGACCCCTCTTTTGATGCCCTCAGCAGTGAAGAGGAGGAAGAGG 25057
QY 61 AGGAGGAGGAGGATGAAGATGAAGAGAGAGGAAGTGGCAGCTTGGAGCGGCCCGCCCAAGAT 120
Db 25056 AGGAGGAGGAGGATGAAGATGAAGAGAGAGGAAGTGGCAGCTTGGAGCGGCCCGCCCAAGAT 24997
QY 121 GGAGTCAGCTGGGAACCTCCACGCGCCCGCCCTTCCCGCCCGCCCTCATCGAAAAACCT 180
Db 24996 GGAGTCAGCTGGGAACCTCCACGCGCCCGCCCTTCCCGCCCGCCCTCATCGAAAAACCT 24937

QY 181 GCTCACAGCGCCCGCCCGAGCCATGAGAGCCCTCCGGATGCTGCTTACTCTAAAAAGCA 240
Db 24936 GCTCACAGCGCCCGCCCGAGCCATGAGAGCCCTCCGGATGCTGCTTACTCTAAAAAGCA 24877
QY 241 CCTCGCTGACATTCACACTGGAAGCTTTTGGGGGGCCACCGGGCCCGCGGGGCTCG 300
Db 24876 CCTCGCTGACATTCACACTGGAAGCTTTTGGGGGGCCACCGGGCCCGCGGGGCTCG 24817
QY 301 CACACCCCAAGAACCATCTTTTACCCACAGCAAGGGGTGGAGCCACAGGTGCCATCCC 360
Db 24816 CACACCCCAAGAACCATCTTTTACCCACAGGAAGGGGTGGAGCCACAGGTGCCATCCC 24757
QY 361 CTTGTTGTCTGTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGCTGGTCTCTAGGTG 420
Db 24756 CTTGTTGTCTGTTTGTACTCCCCCGGGGGCCACCTCCACCCCGGCTGGTCTCTAGGTG 24697
QY 421 CTTCTATGCTGAGGATGGGTGAGAGGTCTCTCCACCACTGTCCTTGGGCCCGCCCATGG 480
Db 24696 CTTCTATGCTGAGGATGGGTGAGAGGTCTCTCCACCACTGTCCTTGGGCCCGCCCATGG 24637
QY 481 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCCTCTGGACCCCTGACTCGGG 532
Db 24636 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCCTCTGGACCCCTGACTCGGG 24585
RESULT 10
AC127470 215795 bp DNA linear HTG 17-JUL-2002
LOCUS Pan troglodytes clone RP43-145D13, WORKING DRAFT SEQUENCE, 15
DEFINITION unordered pieces.
AC127470
VERSION AC127470.1 GI:21886866
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE chimpanzee
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 215795)
AUTHORS Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S. M., Benjamin, B., Blakesley, R. W., Bouffard, G. G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N. L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Idol, J. R., Karlins, E., Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q. L., Maduro, V. B., Margulies, E. H., Mastello, C., Maskeri, B., Mastrian, S. D., McCloskey, J. C., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M. E., Prasad, A., Schueler, M. G., Stantropop, S., Thomas, J. W., Thomas, P. J., Touchman, J. W., Tsurgan, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E. D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 215795)
Green, E. D.
Direct Submission
Submitted (17-JUL-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_zoehnhgri.nih.gov
----- Project Information
Center project name: cms
Center clone name: 145D13
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 203929 bases at least Q40
Consensus quality: 206865 bases at least Q30
Consensus quality: 209181 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 214395; sum-of-contigs
Quality coverage: 7.30x in Q20 bases; agarose-fp

Quality coverage: 6.47x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3593: contig of 3593 bp in length
* 3594: gap of unknown length
* 3694 8001: contig of 4308 bp in length
* 8002 8101: gap of unknown length
* 8102 15639: contig of 7538 bp in length
* 15640 15739: gap of unknown length
* 15740 24775: contig of 9036 bp in length
* 24776 24875: gap of unknown length
* 24876 30150: contig of 5275 bp in length
* 30151 30250: gap of unknown length
* 30251 39212: contig of 8962 bp in length
* 39213 39312: gap of unknown length
* 39313 49342: contig of 10030 bp in length
* 49343 49442: gap of unknown length
* 49443 60957: contig of 11514 bp in length
* 60957 74520: gap of unknown length
* 74521 74620: contig of 13464 bp in length
* 74621 89522: contig of 14902 bp in length
* 89523 89623: gap of unknown length
* 89623 105537: contig of 15915 bp in length
* 105538 121197: gap of unknown length
* 105638 121197: contig of 15560 bp in length
* 121198 121297: gap of unknown length
* 121298 138883: contig of 17586 bp in length
* 138884 138983: gap of unknown length
* 138984 166802: contig of 27819 bp in length
* 166803 166903: gap of unknown length
* 166903 215795: contig of 48893 bp in length.

FEATURES
Source

Location/Qualifiers
1. .215795
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-145D13"
/clone_lib="RP43"

misc_feature
1. .3593
/note="assembly_fragment"
misc_feature
3694. .8001
/note="assembly_fragment"
misc_feature
8102. .15639
/note="assembly_fragment"
misc_feature
15740. .24775
/note="assembly_fragment"
misc_feature
24876. .30150
/note="assembly_fragment"
misc_feature
30251. .39212
/note="assembly_fragment"
misc_feature
39313. .49342
/note="assembly_fragment"
misc_feature
49443. .60956
/note="assembly_fragment"
misc_feature
61057. .74520
/note="assembly_fragment"
misc_feature
74621. .89522
/note="assembly_fragment"
misc_feature
89623. .105537
/note="assembly_fragment"
misc_feature
105638. .121197
/note="assembly_fragment"
misc_feature
121298. .138883
/note="assembly_fragment"
misc_feature
138984. .166802
/note="assembly_fragment"

misc_feature 166903. .215795
/note="assembly_fragment"
BASE COUNT 52160 a 54859 c 53752 g 53513 t 1511 others
ORIGIN
Query Match 26.6%; Score 530.4; DB 2; Length 215795;
Best Local Similarity 99.8%; Pred. No. 2.3e-119;
Matches 531; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCCTCCCTGTCCTCCCGACCCCTCTTTGATGCTCAGCAAGTGAAGAGAGGAGGAGG 60
|||||
Db 205638 GCCTCCCTGTCCTCCCGACCCCTCTTTGATGCTCAGCAAGTGAAGAGGAGGAGGAGG 205697
Qy 61 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGGAGTGGAGCTTGGAGGTGCCCAAGAT 120
|||||
Db 205698 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGTGGAGCTTGGAGGTGCCCAAGAT 205757
Qy 121 GGAGTCAGCTGGGAACCTCCAGGGGCCCGCCCTTCCTCCGCCCACTCATCGAAAAA 180
|||||
Db 205758 GGAGTCAGCTGGGAACCTCCAGGGGCCCGCCCTTCCTCCGCCCACTCATCGAAAAA 205817
Qy 181 GCTCACAGCGCGCGCGAGCCATGAGAGCTTCCGGATGCTCTACTCAAAAAACA 240
|||||
Db 205818 GCTCACAGCGCGCGCGAGCCATGAGAGCTTCCGGATGCTCTACTCAAAAAACA 205877
Qy 241 CTTGCTGACATTCACATGGAAGCTTTGGGGCGCGCCACCGGGCGCGGGGGGCGCTCG 300
|||||
Db 205878 CTTGCTGACATTCACATGGAAGCTTTGGGGCGCGCCACCGGGCGCGGGGGCGCTCG 205937
Qy 301 CACACCCCAAGAACATCTTTCACCCAGCAAGGGGGTGCAGCCACAGGTGCCATCCC 360
Db 205938 CACACCCCAAGAACATCTTTCACCCAGCAAGGGGGTGCAGCCACAGGTGCCATCCC 205997
Qy 361 CTTGTTGCTGTTTACCTCCCGGGGGCCACCTCCACCGCGGTGGTCTGCTAGTGG 420
Db 205998 CTTGTTGCTGTTTACCTCCCGGGGGCCACCTCCACCGCGGTGGTCTGCTAGTGG 206057
Qy 421 CTCTCATGGCTGAGGATGGGTGAGAGGTCTCCACCAAGTGCCTTGGGGGGGGGGGGGG 480
Db 206058 CTCTCATGGCTGAGGATGGGTGAGAGGTCTCCACCAAGTGCCTTGGGGGGGGGGGGGG 206117
Qy 481 AGGAAGATGGACTAGGTGGACTCCAAAGTCTCTCTCGACCCCTGACTCGGG 532
Db 206118 AGGAAGATGGACTAGGTGGACTCCAAAGTCTCTCTCGACCCCTGACTCGGG 206169
RESULT 11
AC127468 173588 bp DNA linear HTG 17-JUL-2002
LOCUS
DEFINITION
Papio cynocephalus anubis clone RP41-263F8, WORKING DRAFT SEQUENCE,
13 unordered pieces.
AC127468
ACCESSION
AC127468.1 GI:21886864
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
olive baboon.
SOURCE
ORGANISM
Papio cynocephalus anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Papio.
REFERENCE
1 (bases 1 to 173588)
AUTHORS
Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C.,
Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J.,
Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P.,
Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,
Margulies, E.H., Masiello, C., Maskeri, B., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Paquirigan, C., Pearson, R.,
Portnoy, M.E., Prasad, A., Schueler, M.G., Stantipop, S., Thomas, J.W.,
Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Wecherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
REFERENCE
2 (bases 1 to 173588)

[illegible]

Db 136003 GGGGTGGCTACAGAAACCCCTTCTTCCTCTGTTTCAGGGGAGTGTGCCCTGTGG 136062

QY 1787 CTTGGGTGAGCAGTCTATCCCTCCCTCCCGTGCAGGAGCAGGAAATCAGTCTGGG 1846

Db 136063 CTTGGGTGAGCAGTCTATCCCTCCCTCCCGTGCAGGAGCAGGAAATCAGTCTGGG 136122

QY 1847 GTGTGGCGGCAATAGATCACTGCTCCAGATCTTCAACATTTTATATATATA 1906

Db 136123 GGTGTGGCGGCAATAGATCACTGCTCCAGATCTTCAACATTTT-TTTTTTTTA 136181

QY 1907 TATATATATATATATAT 1925

Db 136182 TATATATATATATATAAT 136200

RESULT 15

AC126925/c

LOCUS

DEFINITION

Canis familiaris clone RP81-332E11, WORKING DRAFT SEQUENCE, 15

unordered pieces.

AC126925

AC126925.1 GI:21724102

HTG: HTGS_PHASE1; HTGS_DRAFT.

dog.

ORGANISM

Canis familiaris

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

1 (bases 1 to 161428)

Akter.N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,

Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,

Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,

Hadighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,

Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,

Marquies,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,

McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,

Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,

Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,

Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

NISC Comparative Sequencing Initiative

unpublished

2 (bases 1 to 161428)

Green,E.D.

Direct Submission

Submitted (10-JUL-2002) NIH Intramural Sequencing Center, 8717

Grovemont Circle, Gaithersburg, MD 20877, USA

----- Genome Center Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc.zoo@nhgri.nih.gov

----- Project Information

Center project name: cwp

Center clone name: 332E11

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 148712 bases at least Q40

Consensus quality: 151471 bases at least Q30

Consensus quality: 153125 bases at least Q20

Insert size: 152000; agarose-fp

Insert size: 160028; sum-of-contigs

Quality coverage: 5.36x in Q20 bases; agarose-fp

Quality coverage: 5.10x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 15 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 4734: contig of 4734 bp in length

* 4835: gap of unknown length

* 7325: contig of 2491 bp in length

* 7425: gap of unknown length

* 7426: contig of 4836 bp in length

* 12261: gap of unknown length

* 12362: contig of 6670 bp in length

* 19131: gap of unknown length

* 27849: contig of 8718 bp in length

* 27850: gap of unknown length

* 31801: contig of 3852 bp in length

* 31802: gap of unknown length

* 40781: contig of 8880 bp in length

* 40782: gap of unknown length

* 40783: contig of 8902 bp in length

* 49883: gap of unknown length

* 59150: contig of 9267 bp in length

* 59151: gap of unknown length

* 68679: contig of 9429 bp in length

* 68680: gap of unknown length

* 78984: contig of 10205 bp in length

* 78985: gap of unknown length

* 93446: contig of 14362 bp in length

* 93447: gap of unknown length

* 102175: contig of 8629 bp in length

* 102176: gap of unknown length

* 131685: contig of 29410 bp in length

* 131686: gap of unknown length

* 131786: contig of 29643 bp in length.

Location/Qualifiers

1..161428

/organism="Canis familiaris"

/db_xref="taxon:9615"

/clone_lib="RP81-332E11"

/clone_lib="RP81"

1..4734

/note="assembly_fragment"

4835..7325

/note="assembly_fragment"

7426..12261

/note="assembly_fragment"

12362..19031

/note="assembly_fragment"

19132..27849

/note="assembly_fragment"

27850..31801

/note="assembly_fragment"

31802..40781

/note="assembly_fragment"

40882..49783

/note="assembly_fragment"

49884..59150

/note="assembly_fragment"

59251..68679

/note="assembly_fragment"

68780..78984

/note="assembly_fragment"

79085..93446

/note="assembly_fragment"

93547..102175

/note="assembly_fragment"

clone_end:SP6

vector_side:right

102276..131685

/note="assembly_fragment"

131786..161428

/note="assembly_fragment"

BASE COUNT 39586 a 41863 c 41233 g 37340 t 1406 others

ORIGIN

Query Match 23.8%; Score 474.4; DB 2; Length 161428;

Best Local Similarity 93.2%; Pred. NO. 1.3e-105;

Matches 496; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1 GCCTCCCTGTCCCGACCCCTCTTTTGTATGCCTCAGCAAGTGAAGAGAGAGAGAGAGG 60
|||||
Db 88615 GCCTCCCTGTCCCTCGGCCCTCTTTTGTATGCCTCAGCTAGTGAAGAGAGAGAGAGAGG 88556
Qy 61 AGGAGGAGGAGGATGAAGATGAAGAGAGAGAGAGAGAGTGGCAGCTTGGAGGCTGCCCCCAAGAT 120
|||||
Db 88555 AGGAAGATGAGGAGAGATGAGGAAGAGAGAGAGTGGCAGCTTGGAGGCTGCCCTCAAGAT 88496
Qy 121 GGAGTCAAGCTGGGAACCTCCAGCGGCCCGCCCTTCCCGCCCACTCATCGAAAAACCT 180
|||||
Db 88495 GGGGTCAAGCTGGGGCCCTCCAGCGACCTCGCCCTCCCGCCCTACTCATCGAAAAACCT 88436
Qy 181 GCTCACAGCGCCCGCCGAGCCATGAGAGCTTCCGGATGCTGCTCTACTCAAAAAAGCA 240
|||||
Db 88435 GCTCACAGCGCCCGCCGAGCCCAAGAGAGCTTTCGGATGCTGTTCTACCCAAAAAGCA 88376
Qy 241 CCTCGCTGACATTCACATGGAAGCTTTGGGGGCCACCGGGGCCGGGGGGCCCTCG 300
|||||
Db 88375 CCTCGTTGACATTCACATGGAAGCTTTGGGGGCCACCGGGGCCGGGGGCCGAGCCTCG 88316
Qy 301 CACACCCCAAGAACATCTTTTCAACCCAGCAAGGGGTGGCGACGCCACAGGTGCCATCCC 360
|||||
Db 88315 CACACCCCAAGAACATCTTTTCAACCCAGCAAGGGGTGGCGACGCCACAGGTGCCATCTC 88256
Qy 361 CCTGTTGCTGTTTGTACTCCCGCCGGGGCCACCTCCACCCCGGCTGGGTCTGCTAGGTG 420
|||||
Db 88255 CCTGCTGCTGTTTGTACTCCCGCCGGGGCCACCTCCACCCCGGCTGGGTCTGCTAGGTG 88196
Qy 421 CTCTCATGGCTGAGATGGGTGAGAGGGTCTCCACAGTCCCTCTGGGGCCCCCATGG 480
|||||
Db 88195 CTCTCATGGCTGAGATGGGTGAGAGGGTCTCCACCAATGTCCTCTGGGGCCCCAGTGG 88136
Qy 481 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCCTCTGGACCTGACTCGGG 532
|||||
Db 88135 AGGAAGATGGATTAAGGTGACTCCAAAGTCTCCTCTGGATCTGACTCTGG 88084

Search completed: December 16, 2002, 17:29:13
Job time : 6622 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2002, 21:18:11 ; Search time 2422 Seconds
(without alignments)
2220.025 Million cell updates/sec

Title: US-09-848-852A-3
Perfect score: 1772
Sequence: 1 MYSAQRFWGTIARRGAHLA.....DMPKLRRIQYKELCHCKLTV 332

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=fragnet_p2n_model -DEV=xlp
-Q=/cqn2_1/USPTO_spool/US09848852/runat_16122002_132040_19159/app_query.fasta_1.519
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09848852 @CGN 1.1 2024 @runat_16122002_132040_19159 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1391	78.5	1071	14	BM806659	BM806659 AGENCOURT
2	1380	77.9	1060	14	BM806502	BM806502 AGENCOURT
3	1250	70.5	804	12	BG870520	BG870520 602791439
4	1172.5	66.2	728	12	BG865703	BG865703 602783877
5	1133	63.9	985	13	BI456217	BI456217 603172996
6	1097	61.9	626	14	BM696661	BM696661 UI-E-DW0-
7	1073	60.6	905	14	BQ898865	BQ898865 AGENCOURT
8	985.5	55.6	1134	14	BM804530	BM804530 AGENCOURT
c 9	979	55.2	667	17	AZ399027	AZ399027 IMO164L10
c 10	975	55.0	616	10	AW916790	AW916790 EST348198
11	951	53.7	620	10	AW578424	AW578424 RCI-CT024
12	951	53.7	620	10	AW604359	AW604359 RCI-CT024
13	929	52.4	572	10	BE077141	BE077141 RCI-BT060
c 14	924	52.1	626	9	AL578220	AL578220 AL578220
15	913	51.5	870	13	BI152305	BI152305 602917742
16	903	51.0	549	13	BM507574	BM507574 ih4le03.y
17	877	49.5	593	10	AW643059	AW643059 cm25f04.w
18	865	48.8	819	12	BF139554	BF139554 601785671
19	852	48.1	545	17	AQ978128	AQ978128 RPCI-23-3
20	834	47.1	504	12	BF954315	BF954315 QV2-NN004
21	833	47.0	621	13	BI392544	BI392544 pcpin.pk0
22	818	46.2	477	10	AW825381	AW825381 ue14e04.y
c 23	812	45.8	817	9	AL581080	AL581080 AL581080
24	809	45.7	489	12	BF458035	BF458035 UI-M-BZ1-
25	804	45.4	458	14	W63893	W63893 md85h02.r1
c 26	801	45.2	835	13	BI732013	BI732013 603355881
27	798	45.0	449	9	AA549588	AA549588 vk76e01.s
28	798	45.0	689	12	BF608506	BF608506 MY1_00144
29	794	44.8	474	14	BM821282	BM821282 K-EST0090
30	794	44.8	605	14	BM829799	BM829799 K-EST0102
c 31	794	44.8	1070	13	BI247525	BI247525 602960160
32	789	44.5	1828	11	BC000958	BC000958 Homo sapi
33	784	44.2	1836	11	AF335474	AF335474 Homo sapi
34	782	44.1	447	9	AI148063	AI148063 qg61g08.r
35	782	44.1	468	9	AI261629	AI261629 qz30e09.x
c 36	782	44.1	478	14	BM941363	BM941363 UI-M-BZ1-
c 37	776	43.8	599	13	BI017569	BI017569 PK3-ET027
38	770	43.5	1099	13	BM55097	BM55097 AGENCOURT
39	768	43.3	632	12	BG083248	BG083248 H3086807-
c 40	765	43.2	650	14	BQ261255	BQ261255 fz71g09.x
c 41	759	42.8	665	13	BM185366	BM185366 fv46c09.x
c 42	753	42.5	1158	14	BQ934444	BQ934444 AGENCOURT
43	749	42.3	595	14	BQ395166	BQ395166 NISC_rg13
c 44	742	41.9	676	10	AW961869	AW961869 EST373942
45	738	41.6	499	12	BF443604	BF443604 261154 MA

ALIGNMENTS

RESULT 1
BM806659
LOCUS
DEFINITION BM806659 1071 bp mRNA linear EST 05-MAR-2002
AGENCOURT_6542888 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5548958
5', mRNA sequence.
ACCESSION BM806659
VERSION BM806659.1 GI:19123482
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1071)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLML2258 row: 1 column: 15
 High quality sequence stop: 648.
 Location/Qualifiers
 1..1071

FEATURES
source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5548958"
 /clone_lib="NIH_MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site.1: NotI;
 Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."
 255 a 264 c 271 g 278 t 3 others

BASE COUNT
ORIGIN

Alignment Scores:
 Pred. No.: 2.28e-169 Length: 1071
 Score: 1391.00 Matches: 264
 Percent Similarity: 99.25% Conservative: 0
 Best Local Similarity: 99.25% Mismatches: 1
 Query Match: 78.50% Indels: 1
 DB: 14 Gaps: 0

US-09-848-852a-3 (1-332) x BM806659 (1-1071)

QY 68 ArgGluGluHisValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGly 87
 DB 2 CGAGAGGAGCATGTGACCTGGTACAGAGCATCTTGGACGAATCTCTCAACGATATGC 61
 QY 88 SerLeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGln 107
 DB 62 AGCCTCATACCCCTCAGCACTGATGAGTAGTAGAGAAAGCTGGAGGACATTTTCCAGCAG 121
 QY 108 GluPheSerThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArg 127
 DB 122 GAGTTTCCACCCCTCCAGGAAGGCGCTGTGTTCAGCTGATCCAGTCTTACCAAGCGG 181
 QY 128 MetProGlyAsnAlaMetValArgGlyPheArgValAlaIleTyrLysArgHisValLeuThr 147
 DB 182 ATGCCAGGCAATGCCATGTTGAGGGCTTCCGAGTGGCTTATAAGCGGCACGCTGTGACC 241
 QY 148 MetAspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMet 167
 DB 242 ATGGATGACTTGGGGACCTGTGTGACACAGACTGGCTCAATGACCAAGGTGATGAACATG 301
 QY 168 TyrGlyAspLeuValMetAspThrValProGluLysValHisPheAsnSerPhePhe 187
 DB 302 TATGGAGACCTGTGTCATGACACAGTCCCTGAAAGAGTGCATTTCTCAATAGTTTCTTC 361
 QY 188 TyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgThrPheLysAsnValAsp 207
 DB 362 TATGATAAATCCGTACCAAGGGTTATGATGGGTCAAAAGGTGGACCAAAACGTGGAC 421
 QY 208 IlePheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIle 227
 DB 422 ATCTTCAATAAGAGAGCTACTGTCTAAATCCCATCCACTGGAGGTGCATTTGGTCCCTCATC 481
 QY 228 SerValAspValArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArg 247
 DB 482 TCTGTTGATGTAGGGCGACGACCATCACCTATTTTGGACTCGCAGCGTACCCCTAAACCCG 541

QY 248 ArgCysProLysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeu 267
 DB 542 CGTGCCTTAAGCATATTGCCAAGTATCTACAGCGAGCGGTAAAGAACCGACTG 601
 QY 268 AspPheHisGlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAsnAsp 287
 DB 602 GATTTCCACCGAGGCTGGAAAGGTTACTTCAAAATGAATGTGGCCAGGAGAGATAATGAC 661
 QY 288 SerAspCysGlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPhe 307
 DB 662 AGTCACTGTGGTCTTTTGTGTTCAGTACTGCAACATCTGGCCCTGTCTCAGCCATTC 721
 QY 308 SerPheThrGlnGlnAspMetProLysLeuArgArgGlnIleTyrLysGluLeuCysHis 327
 DB 722 AGCTTCACCCAGCAGGACATGCCAAACTTCGTGCGCAGATCTFACAAGGAGCTGTGTCTCAC 781
 QY 328 Cys-LysLeuThrVal 332
 DB 782 TGGCAAACTCACTGTG 797

RESULT 2
BM806502

LOCUS
 DEFINITION BM806502 1060 bp mRNA linear EST 05-MAR-2002
 5', mRNA sequence.
 ACCESSION BM806502 GI:19123325
 VERSION
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1060)
 NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLML2258 row: h column: 18

High quality sequence stop: 652.

FEATURES

source

Location/Qualifiers
 1..1060
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5548958"
 /clone_lib="NIH_MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site.1: NotI;
 Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."
 BASE COUNT 248 a 272 c 304 g 236 t
 ORIGIN

Alignment Scores:

Pred. No.: 6.01e-168 Length: 1060
 Score: 1380.00 Matches: 285
 Percent Similarity: 95.35% Conservative: 2
 Best Local Similarity: 94.68% Mismatches: 6
 Query Match: 77.88% Indels: 9
 DB: 14 Gaps: 2

US-09-848-852a-3 (1-332) x BM806502 (1-1060)

QY 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAlaHis-LeuAl 20

```
|||||
Db 40 ATGTACTCTGCCAACGGTTTGGGGCAAACTCTGGCCAGAGGGAGCGCAGCTTGGC 99
QY 20 aProProaspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaGI 40
Db 100 ACCCCCTGATGCCAGCATCTCATCAGCAATGTGTCAGCATCGGGGACCATGTGGCCCA 159
QY 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluAraGpGroglyGluLy 60
Db 160 GGAGCTTTTCAGGGCTCAGATTGGGCATGGCAGAGAGGAGAGGCGCTGGGGAGNA 219
QY 60 salaGlyGlnHisSerProLeuAraGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 220 ACCCGGCCAGCACAGCCCTCGAGAGAGGAGCATGTGACCTCGTACAGCATCTTGA 279
QY 80 pGluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValValGluLy 100
Db 280 CGAATTCCTCAACAGTATGGCAGCTCATACCCCTCAGCACTGATGAGGTAGTAGAA 339
QY 100 sLeuGluAspIlePheGlnGlnPheSerThrProSerArgLysGlyLeuValLeuGI 120
Db 340 GCTGGAGGACATTTCCAGCAGAGTTTCCACCCCTCCAGAGAGGCGCTGGTGTGA 399
QY 120 nLeuIleGlnSerTyrglnArqMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 400 GCTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGTGTGAGGGGCTCCGAGTGC 459
QY 140 aTyrglyArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrglyGlnAsnTrpLe 160
Db 460 TTATAAGGGGACGGTGTGACCATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCT 519
QY 160 uAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrValProGluLysVa 180
Db 520 CAATCACCAGGTGATGACATGTATGGAGACCTGTGTCATGGACACAGTCCCTGAAAAGT 579
QY 180 lHisPheAsnSerPhePheTyrglyAspLysLeuArgThrLysGlyTyrglyVally 200
Db 580 GCATTTCTTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGTGAA 639
QY 200 sArgThrPheLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisle 220
Db 640 AAGGTGGACCAAAACGTGGGACATCTCAATAAGGAGTACTGCTAAAT-CCCATCCACT 698
QY 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrglyPheAs 240
Db 699 GGAGGTGATTTGGTCCCTCATCTCTGTGATGTGAGCGACGCCACCATCTATTGA 758
QY 240 pSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrglyLeuGlnAlaGI 260
Db 759 CTCGAGCGGTACCTAGACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAGGCAGA 818
QY 260 u-AlaValLys-LysAspArgLeu-AspPheHisGlnGly---TrpLysGlyTyrgly 278
Db 819 AGCGCGTTAAGAAACACCGACTGGGATTTCCACCACGGGCTGGGAAAGGGTTACTTCAA 878
QY 278 sMetAsn-ValAlaArgGlnAsn-AsnAspSerAsp---CysGlyAlaPhe 293
Db 879 AATGAATGGTGCCAGGCAAGATAATGACACTGACCGGTGGGGGCTTTT 929

RESULT 3
BG870520
LOCUS 602791439F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4922696 5',
DEFINITION BG870520 804 bp mRNA linear EST 29-MAY-2001
ACCESSION BG870520
VERSION BG870520.1 GI:14221060
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
```

TITLE JOURNAL COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILLNL at:
http://image.llnl.gov
Plate: LLAM10842 row: n column: 09
High quality sequence stop: 789.

FEATURES
source
1. 804
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4922696"
/clone_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1;
NotI; Site: 2; SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 223 a 195 c 202 g 184 t
ORIGIN

Alignment Scores:
Pred. No.: 2,99e-151 Length: 804
Score: 1250.00 Matches: 235
Percent Similarity: 98.74% Conservative: 1
Best Local Similarity: 98.33% Mismatches: 2
Query Match: 70.54% Indels: 1
DB: 12 Gaps: 0

US-09-848-852A-3 (1-332) x BG870520 (1-804)

QY 95 AspGluValValGluLysLeuGluAspIlePheGlnGlnPheSerThrProSerArg 114
Db 38 GATGAAGTTGTAGAGAAGTTGGAGGACATTTTCCAGCAGGAGTTCTTACACCCCTCAAGG 97
QY 115 LysGlyLeuValLeuGlnLeuIleGlnSerTyrglnArqMetProGlyAsnAlaMetVal 134
Db 98 AAGAGTCTGGTACTACAGCTGATCCAGTCTTATCAGCGGATGCCAGCAACCTATGGTA 157
QY 135 ArgGlyPheArgValAlaTyrglyArgHisValLeuThrMetAspAspLeuGlyThrLeu 154
Db 158 AGGGGCTTCCGGGTATCTCTATTAAGCGACACGCTGCCTCACCATGGATGACTTGGGTACCTTA 217
QY 155 TyrglyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAsp 174
Db 218 TATGGACAGAACTGCTCAATCACCAGGTGATGAACATGTATGGAGATCTGGTCATGGAC 277
QY 175 ThrValProGluLysValHisPheAsnSerPhePheTyrglyAspLysLeuArgThrLys 194
Db 278 ACAGTCCCTGAAAAGTGCATTTCTCAACAGCTTCTTCTATGATAAACTCCGTACCAAG 337
QY 195 GlyTyrglyValLys-ArgThrThrLysAsnValAspIlePheAsnLysGluLeuLeu 214
Db 338 GGTATGATGGGGTAAACGAGGTGGACCAAAAATGTGGACATCTTCAATAAGGAATTAAT 397
QY 214 uLeuIleProIleHisLeuGluValHisTrpSerLeuIleSerValAspValArgArg 234
Db 398 GCTAATCCCATCCATCTGGAGGTGCATGGTCCCTTATCTCAGTTGATGTAAAGCGCAGC 457
QY 234 gThrIleThrTyrglyAspSerGlnArgThrLeuAsnArgArgCysProLysHisIleAl 254
Db 458 TACCATCATCTATTGATCCAGGAACTTAATCCCGCTGCCCTTAACATATGCG 517
QY 254 aLysTyrglyGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLy 274
Db 518 CAAGTATCTACAGGCAGAGGCAGTCAAAAACACCGACTGGGACTTCCATCAGGCGCTGAA 577

```

QY 274 sGlyTyrPhelYsnMetAsnValAlaArgGlnAsnAspSerAspCysGlyAlaPheVa 294
|||||
Db 578 AGGTTACTTCAAAATGATGTGCCAGGAGATAATGACAGTGTGGTGGCTTTGT 637

QY 294 lLeuGlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMe 314
|||||
Db 638 ATTAACAGTACTGCAAGCACCTGGCCCTGTCTCAGCCATTTCAGCTTCCAGCAGGAGACAT 697

QY 314 tProLysLeuArgArgGlnIleYrLysGluLeuCysHisCysLysLeuThrVal 332
|||||
Db 698 GCCAAACTCCGTGTGCAGATCTACAGGAAGTGTGTCTACTGCAAACTACTGTG 752

RESULT 4
BG865703 728 bp mRNA linear EST 29-MAY-2001
LOCUS 602783877F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4909938 5',
DEFINITION mRNA sequence.
ACCESSION BG865703
VERSION BG865703.1 GI:14216243
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 728)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10809 row: j column: 19
High quality sequence stop: 728.
FEATURES
source
1. .728
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4909938"
/lab_host="NCI_CGAP_SG2"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 198 a 168 c 186 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 2 99e-141 Length: 728
Score: 1172.50 Matches: 234
Percent Similarity: 97.53% Conservative: 3
Best Local Similarity: 96.30% Mismatches: 6
Query Match: 66.17% Indels: 3
DB: 12 Gaps: 0

US-09-848-852A-3 (1-332) x BG865703 (1-728)

QY 29 SerAsnValCysSerTleGlyAspHisValAlaGlnGluLeuPheGlnGlySerAspLeu 48
|||||
Db 2 AGTAATGTGTGCAGCATGTGAGACCACGTCGGCTCAGGAACCTTTTCAGAGCTCTGACTTG 61

QY 49 GlyMetAlaGluAlaGluArgProGlyGluLysAlaGlyGlnHisSerProLeuArg 68
|||||
Db 62 GGCATGCA-GAAGAGGCAGATCGGATGGGGGAAAAGCTGGCCAGCATAGCCCCCTTCGG 120

```

```

QY 69 GluGluHisValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySer 88
|||||
Db 121 GAGAACATGTGACCTGTGTGCAGAGATATCTTAGATGAATTCCTTCAAACTTATGGCAGC 180

QY 89 LeuIleProLeuSerThrAspGluValValGluLysLeuGluAspIlePheGlnGlnGlu 108
|||||
Db 181 CTCATCCCTCTCAGCACTGATGAAGTTGTAGAGAAGTTGGAGGACATTTTCCAGCAGGAG 240

QY 109 PheSerThrProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMet 128
|||||
Db 241 TTCTCTACACCCCTCAAGGAAGAGTCGTGGTACTACAGCTGATCCAGTCTTATCAGCGGATG 300

QY 129 ProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThrMet 148
|||||
Db 301 CCAGGCAACGCTATGTAAGGGCTTCCGGGTATCTCTATAAGCGACACGCTGCTCACCATG 360

QY 149 AspAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyr 168
|||||
Db 361 GATGACTTGGGTACCTTATATGACAGCAACTGGCTCAATGACCAGGTGATGAACATGTAT 420

QY 169 GlyAspLeuValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyr 188
|||||
Db 421 GGAGATCTGGTCTATGACACAGCTCCCTGAAAGGTGCATTTCTTCAACAGCTTCTTCTAT 480

QY 189 AspLysLeuArgThrLysGlyTyrAspGlyValLysArgTyrThrLysAsnValAspIle 208
|||||
Db 481 GATAAACTCCGTACCAAGGGTTATGATGGGTAAAGAGGTGGACCAAAAATGTGGACATC 540

QY 209 PheAsnLysGluLeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIleSer 228
|||||
Db 541 TTCATAAAGGAATTACTTGGTAATCCCATCCCATCTGGAGTGCACCTGGTCCCTTATCTCA 600

QY 229 ValAspValArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgArg 248
|||||
Db 601 GTTGATGTAAAGCGAGCTACCATCACCATTATTTTGACTCCCGAGGAACTCTTAAATGCCCG 660

QY 249 CysProLysHisIleAlaLysTyrLeuGlnAlaGluAlaValLysLysAspArgLeuAsp 268
|||||
Db 661 TGCCT-AAGCATATTGC-AAGTATCTACAGGAGAGGAGTCAGTCAAAAAGACCGACTGGAC 718

QY 269 PheHisGln 271
|||||
Db 719 TTCCATCAG 727

RESULT 5
BI456217 985 bp mRNA linear EST 21-AUG-2001
LOCUS 603172996F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5252591 5',
DEFINITION mRNA sequence.
ACCESSION BI456217
VERSION BI456217.1 GI:15246873
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 985)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1637 row: o column: 24
High quality sequence stop: 746.
FEATURES
source
1. .985

```

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="5252591"
 /clone_lib="NCI_OGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 263 a 220 c 243 g 258 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 6,62e-136 Length: 985
 Score: 1133.00 Matches: 208
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.52% Mismatches: 0
 Query Match: 63.94% Indels: 0
 DB: 13 Gaps: 0

US-09-848-852A-3 (1-332) x B1456217 (1-985)

QY 124 SerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArg 143
 DB 2 TCTATACGGGATGCCAGCAACGCTATGGTAAGGGGCTTCGGGTATCCTATAAGCGA 61
 QY 144 HisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGln 163
 DB 62 CACGTGCTCACCATGATGCTGGGTACCTTATATGGAGAACTGGCTCAATGACCAG 121
 QY 164 ValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysValHisPhePhe 183
 DB 122 GTGATGAACATGATGAGATCTGGTCATGGACACAGTCCCTGAAAAGGTGCATTTCTTC 181
 QY 184 AsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTTPThr 203
 DB 182 AACAGCTTCTCTATGATTAACCTCCGTACCAAGGTTATGATGGGTAAAGAGTGGACC 241
 QY 204 LysAsnValAspTlePheAsnLysGluLeuLeuLeuLeuLeuLeuLeuLeuLeuValHis 223
 DB 242 AAAAAATGTGGACATCTTCAATAAGGAATTACTGCTAATCCCATCCATCTGGAGGTGCAC 301
 QY 224 TrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAspSerGlnArg 243
 DB 302 TGGTCCCTTATCTCACTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 361
 QY 244 ThrLeuAsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaGluAlaValLys 263
 DB 362 ACTCTAAATCGCGCTGCCCTAGCATATATGCCAAGTATCTACAGCAGGAGCGATCAAA 421
 QY 264 LysAspArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArg 283
 DB 422 AAAGACCGAGTGGACTTCCATCAGGCTGGAAAGGTTACTTCAAAATGAATGTGGCCAGG 481
 QY 284 GlnAsnAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeu 303
 DB 482 CAGAAATAAATGACAGTACACTGTGGTGCCTTTGTATTACAGTACTGCAAGCACCTGGCCCTG 541
 QY 304 SerGlnProPheSerPheThrGlnGlnAspMetProLysLeuArgArgGlnIleTyrLys 323
 DB 542 TCTCAGCCATTCAGCTTACCCACAGGAGACATGCCCAAACTCGCTGATCTACAAAG 601
 QY 324 GluLeuCysHisCysLysLeuThrVal 332
 DB 602 GAACCTGTGCTCACTGCAAACTCACTGTG 628

RESULT 6

BM696661

LOCUS

626 bp mRNA linear EST 28-FEB-2002

DEFINITION

UI-E-DW0-agk-m-22-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
 UI-E-DW0-agk-m-22-0-UI 5', mRNA sequence.

ACCESSION BM696661
 VERSION BM696661.1
 KEYWORDS GI:19009919
 EST.

SOURCE

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 626)

AUTHORS

Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene

JOURNAL

MEDLINE

COMMENT

Genome Res. 6 (9), 791-806 (1996)
 97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..626

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UI-E-DW0-agk-m-22-0-UI"

/tissue_lib="UI-E-DW0"

/tissue_type="adult"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a

modified polylinker; Site_1: EcoR I; Site_2: Not I;

UI-E-DW0 is a cDNA library containing the following

tissue(s): lens. The library was constructed according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded

cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pT7T3-Pac vector. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CGATTAGCGA. This library

was created for the program, Gene Discovery in the Visual

System, supported by National Eye Institute (NEI)."

BASE COUNT 159 a 157 c 172 g 138 t

ORIGIN

Alignment Scores:

Pred. No.: 1,51e-131 Length: 626

Score: 1097.00 Matches: 208

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 61.91% Indels: 0

DB: 14 Gaps: 0

US-09-848-852A-3 (1-332) x BM696661 (1-626)

QY 52 GluGluAlaGluArgProGlyGluLysAlaGlyGlnHisSerProLeuArgGluHis 71

DB 2 GAAGAGCAGAGAGCGCTGGGAGAAAGCGCCACACAGCCCTCGCAGAGGAGCAT 61

QY 72 ValThrCysValGlnSerIleLeuAspGluPheLeuGlnThrTyrGlySerLeuPro 91

DB 62 GTGACCTGCTACAGAGCATCTTGGACGAATTCCTTCAAACGTATGGCAGCTCATACC 121

Qy 92 LeuSerThrAspGluValGluLysLeuGluAspIlePheGlnGlnGluPheSerThr 111
 |||||
 Db 122 CTCAGCACTGATGAGTAGTAGAGACAGCTGGAGGACATTTTCCAGCAGGAGTTTCCACC 181
 |||||
 Qy 112 ProSerArgLysGlyLeuValLeuGlnLeuIleGlnSerTyrGlnArgMetProGlyAsn 131
 |||||
 Db 182 CTTTCCAGGAAGGCGCTGGTTCAGCTGATCCAGTCTTACCAGCGGATGCCAGGCAAT 241
 |||||
 Qy 132 AlaMetValArgGlyPheArgValAlaTyrLysArgHisValLeuThrMetAspLeu 151
 |||||
 Db 242 GCCATGGTAGGGCTCCGAGTGGCTTATAAGCGGACGCTGGCTGACCATGGATGACTTG 301
 |||||
 Qy 152 GlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrGlyAspLeu 171
 |||||
 Db 302 GGGACCTTGATGACAGACAGCTGCTCAATGACAGCTGATGAACATGTATGGAGACTG 361
 |||||
 Qy 172 ValMetAspThrValProGluLysValHisPhePheAsnSerPhePheTyrAspLysLeu 191
 |||||
 Db 362 GTCATGGACACAGTCCCTGAAAGGTGCATTTCTTCAATAGTTTCTTCTATGATAAACTC 421
 |||||
 Qy 192 ArgThrLysGlyTyrAspGlyValLysArgTyrThrLysAsnValAspIlePheAsnLys 211
 |||||
 Db 422 CGTACCAGGCTTATGATGGGTGAAAGGTGGACCAAAACGTGGACATCTTCAATAAG 481
 |||||
 Qy 212 GluLeuLeuLeuIleProIleHisLeuGluValHisTrpSerLeuIleSerValAspVal 231
 |||||
 Db 482 GAGCTACTGCTAATCCCATCCACTGGAGTGCATTGGTCCCTCATCTCTGTGTGATGTG 541
 |||||
 Qy 232 ArgArgThrIleThrTyrPheAspSerGlnArgThrLeuAsnArgCysProLys 251
 |||||
 Db 542 AGCGGACGACCATCACCTATTTTGACTCGCAGCGTACCCTAAACCCGCCCTGCCCTAAG 601
 |||||
 Qy 252 HistLeuLysTyrLeuGlnAla 259
 |||||
 Db 602 CATATTGCCAAGTATCTACAGGCA 625
 |||||

RESULT 7
 BQ898865 905 bp mRNA linear EST 16-AUG-2002
 LOCUS
 DEFINITION AGENCOURT_8489704 Lupski_dorsal_root_ganglion Homo sapiens cDNA
 clone IMAGE:6181093 5', mRNA sequence.
 ACCESSION BQ898865
 VERSION BQ898865.1 GI:22290891
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 905)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM13564 row: O column: 14
 High quality sequence stop: 614.
 Location/Qualifiers
 1..905
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Lupski_dorsal_root_ganglion"
 /sex="male"
 /tissue_type="dorsal root ganglia"
 /dev_stage="adult, 36 yr"

FEATURES

source
 BM804530
 LOCUS
 DEFINITION AGENCOURT_6492923 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5554637
 5', mRNA sequence.
 ACCESSION BM804530
 VERSION BM804530.1 GI:19121353
 KEYWORDS EST.
 SOURCE human.
 RESULT 8
 BM804530
 LOCUS
 DEFINITION AGENCOURT_6492923 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5554637
 5', mRNA sequence.
 ACCESSION BM804530
 VERSION BM804530.1 GI:19121353
 KEYWORDS EST.
 SOURCE human.

/lab_host="DH10B"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
 Note: Site_2: SalI: cDNA made by oligo-dT priming.
 Directionally cloned using the following adaptors:
 5'-TCGACCCACGCGTCCG-3' and
 5'-GACTAGTTCTAGATCGAGCGCCGCTT(15)-3'. Size selected >
 1 kb for average insert length 1.7 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine) and is available through Life
 Technologies."

BASE COUNT 212 a 238 c 220 g .235 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3,61e-128 Length: 905
 Score: 1073.00 Matches: 197
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 60.55% Indels: 0
 DB: 14 Gaps: 0

US-09-848-852a-3 (1-332) x BQ898865 (1-905)

Qy 136 GlyPheArgValAlaTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyr 155
 |||||
 Db 2 GGCTTCGAGTGGCTTATAGCGGCAGCTGCTGACCATGGATGACCTGGGGACCTTGTAT 61
 |||||

Qy 156 GlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThr 175
 |||||
 Db 62 GGACAGAACTGGCTCAATGACACAGGTGATGAACATGTATGAGACCTGGTTCATGGACACA 121
 |||||

Qy 176 ValProGluLysValHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGly 195
 |||||
 Db 122 GTCCCTGAAAGGTGCATTTCTTCAATAGTTTCTTATGATAACTCCGTCACAGGGT 181
 |||||

Qy 196 TyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeu 215
 |||||
 Db 182 TATGATGGGTGAAAGGTGGACCAAAACGTGGACATCTTCAATAAGGAGCTACTGCTA 241
 |||||

Qy 216 IleProIleHisLeuGluValHisTrpSerLeuIleSerValAspValArgArgThr 235
 |||||
 Db 242 ATCCCATCCACCTGAGGTGCTATGGTCCCTCATCTCTGTGATGAGCGCAGCACACC 301
 |||||

Qy 236 IleThrTyrPheAspSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLys 255
 |||||
 Db 302 ATACCTATTTTGACTCGCAGCGTACCCTAAACCGCGCTGCCCTAAGCATATATTGCCAAG 361
 |||||

Qy 256 TyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGly 275
 |||||
 Db 362 TATCTACAGCAGCAGCGGTGAAAGAAAGACCGACTGGATTTCCACCGAGGCTGGAAGGT 421
 |||||

Qy 276 TyrPheLysMetAsnValAlaArgGlnAsnAsnAspSerAspCysGlyAlaPheValLeu 295
 |||||
 Db 422 TACTTCAAAATGAATGTGCCAGCGCAATAATGACAGTGACTGTGGTGTGTGTG 481
 |||||

Qy 296 GlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetPro 315
 |||||
 Db 482 CAGTACTGCAGCATCTGGCCCTGTCTCAGCCATTCAGCTTCCACCGAGGACATGCC 541
 |||||

Qy 316 LysLeuArgArgGlnIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
 |||||
 Db 542 AAACCTCGTCGGCAGATCTACAGGAGCTGTGTCACTGCAAACTCACTGTG 592
 |||||

of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 161 a 160 c 160 g 186 t
ORIGIN

Alignment Scores:
Pred. No.: 3.6e-116 Length: 667
Score: 979.00 Matches: 188
Percent Similarity: 92.82% Conservative: 6
Best Local Similarity: 89.95% Mismatches: 9
Query Match: 55.25% Indels: 6
DB: 17 Gaps: 1

US-09-848-852A-3 (1-332) x AZ399027 (1-667)

Qy 126 GlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAlaTyrLysArgHisVal 145
Db 665 CAGCGGATGCCAGACACCGTATGGTAAGGGGCTTCGGGTATCTCTATAAGTGACAGTG 606
Qy 146 LeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeuAsnAspGlnValMet 165
Db 605 CTCACCATGGATGGTGGTACCTTATATGGACAGAACTGGCTCATGCCAGGTGATG 546
Qy 166 AsnMetTyrGlyAspLeuValMetAspThrValProGluLysValHisPheAsnSer 185
Db 545 AACATGTATGGAGACCTGGTCATGGACACAGTCCTGAAAAGGTGCATCTTCTCCACAGT 486
Qy 186 PhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyValLysArgTTrpThrLysAsn 205
Db 485 CTCCTTCATGATAAATCCCGTACCAAGGTTATGATGGGTAAAGAGGTGGACCAAAAT 426
Qy 206 ValAspPheAsnLysGluLeuLeuLeuPheProIleHisLeuGluValHisTrpSer 225
Db 425 GTGGACATCTCAATAGGAATTAATCTCTAATCCCATCCATCTGGAGTGGAGTGTCC 366
Qy 226 LeuIleSerValAspValArgArgArgThrIleThrTyrPheAspSerGlnArgThrLeu 245
Db 365 CTTATCTCAGTTGATGTAAAGCAACATACATCACCTATTCTTACTCCCGCAACTTGA 306
Qy 246 AsnArgArgCysProLysHisIleAlaLysTyrLeuGlnAlaVal-LysLysAs 265
Db 305 AATCGCCGCTGCCCTAAGTATATTTCAAGTATCTACAGGCAGAGGCGAGTCAAAAAAGG 246
Qy 265 pArgLeuAspPheHisGlnGlyTrpLysGlyTyrPheLysMetAsnValAlaArgGlnAs 285
Db 245 CCAAGTGGACTTCATCAGGCTGGAAAGGTATTTCAAATGAATGTGGCCAGGCAGAA 186
Qy 285 nAsnAspSerAspCysGlyAlaPheValLeuGlnTyrCysLysHisLeuAlaLeuSerG1 305
Db 185 TAATGACAGTGACTGTGGTCCCTTGTGTACAGTACTGCAAGCACCTGGCCCTGTCTCA 126
Qy 305 nProPheSerPheThrGlnGlnAsp-MetProLysLeuArgArgGlnIleTyrLysGluL 325
Db 125 GCCATTCAGTTTCACCCAGGAGGACATCT-----CGTCAGATCTGCAAGGAAC 78
Qy 325 euCysHisCysLysLeuThrVal 332
Db 77 TCTGTCTACTGCAAACTCACAGTG 55

RESULT 10

AW916790/c

LOCUS

DEFINITION

AW916790

EST348198 Rat gene index, normalized rat, norvegicus, Bento Soares

AW916790

EST

ACCESSION

VERSION

AW916790.1 GI:8082631

KEYWORDS

SOURCE

616 bp mRNA linear EST 25-MAY-2000
Rattus norvegicus cDNA clone RG1DW31 5' end, mRNA sequence.

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 616)
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat

AUTHORS

Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org

TITLE

This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information
Seq primer: M13 Reverse.
Location/Qualifiers
1..616
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RG1DW31"
Bento Soares
/tissue_type="mix - brain, ovary, placenta, kidney, lung,
liver, embryo, heart, muscle, spleen"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Estimated insert size approx.1 kb"

FEATURES

source

148 a 165 c 153 g 150 t
ORIGIN

Alignment Scores:

Pred. No.: 1.05e-115 Length: 616
Score: 975.00 Matches: 187
Percent Similarity: 95.98% Conservative: 4
Best Local Similarity: 93.97% Mismatches: 7
Query Match: 55.02% Indels: 1
DB: 10 Gaps: 0

US-09-848-852A-3 (1-332) x AW916790 (1-616)

Qy 1 MetTyrSerAlaGlnArgPheTrpGlyThrIleTrpAlaArgGlyAla-HisLeuAl 20
Db 595 ATGTACTCTTACCCTTGTGGGGGACTGTCTGTGGTCTGAAGGGGAGCGTAGCCTGGC 536
Qy 20 aProAspAlaSerIleLeuIleSerAsnValCysSerIleGlyAspHisValAlaG1 40
Db 535 ACCCTCTGATGCCAGCATACTCATCAGCAATGTGTGAGCATTTGGACACCGTGGCTCA 476
Qy 40 nGluLeuPheGlnGlySerAspLeuGlyMetAlaGluAlaGluArgProGlyGluLy 60
Db 475 GGAACCTTTTTCAGAGCTCGGACTTGGGCACCTGCAGAGAGAGCAGATCGCCAGGGAGAA 416
Qy 60 sAlaGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAs 80
Db 415 AGTGGCCAGCATAGCCCCCTGGGGAGGAACATGTGACCTGTGTGCAGAGTATCTTAGA 356
Qy 80 pGluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLy 100
Db 355 TGAATTCCTTCAAACTTATGTAGCTCATCCCTCAGCAGTATGATGAAGTGTAGAGAA 296
Qy 100 sLeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuG1 120
Db 295 GTTGGAGGACATTTCCAGCAGGAGTCTCTACCCCTCCAGGAAGAGTCTGTGTACTACA 236
Qy 120 nLeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAl 140
Db 235 GCTGATCCAATCTTACCAGCGGATGCCAGCAACGCATGGTAAGGGGCTTCGGGGTATC 176
Qy 140 aTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLe 160


```
Db 175 CATAAAGACACGCTGCTCACCATGGATGACTGGTACCTTATATGGACAACTGGCT 116
      :|||||
Qy 160 uAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVa 180
      |||||
Db 115 CAATGACACAGGTGATGAACATATATGGAGATGTTGTTATGGACACAGTCCCTGAAAGGT 56
      |||||
Qy 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGly 198
      |||||
Db 55 GCATTCTTCACAGTTCTTCTATGATAAACTCGTACCAAGGGTTATGATGCC 1
      |||||

RESULT 11
AW578424
LOCUS RC1-CT0249-120100-022-c03 CT0249 Homo sapiens cDNA, mRNA linear EST 16-MAR-2000
DEFINITION AW578424
ACCESSION AW578424
VERSION AW578424.1 GI:7253473
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 620)
JOURNAL HCGP http://www.ludwig.org.br/ORESTES.
COMMENT The FAPESP/LICR Human Cancer Genome Project
Unpublished (1999)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1st2=RC1-CT0249-
120100-022-c03&t3=2000-01-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 583.
High quality Location/Qualifiers
1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0249"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 161 a 146 c 163 g 150 t
ORIGIN

Alignment Scores:
Pred. No.: 1.39e-112 Length: 620
Score: 951.00 Matches: 191
Percent Similarity: 91.04% Conservative: 2
Best Local Similarity: 90.09% Mismatches: 2
Query Match: 53.67% Indels: 17
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x AW578424 (1-620)

Qy 61 AlAGlyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAsp 80
      :|||||
Db 12 TCTGGAGACATCG-----GAC 29
      |||||

Qy 81 GluPheLeuGlnThrTyrGlySerLeuIleProLeuSerThrAspGluValValGluLys 100
      |||||
```

```
Db 30 GAATTCCTTAAACGCTATGGCAGCCTCATATACCCCTCAGCACCTGATGAGGTAGAGAA 89
      |||||
Qy 101 LeuGluAspIlePheGlnGlnGluPheSerThrProSerArgLysGlyLeuValLeuGln 120
      |||||
Db 90 CTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGCCCTGGTGTTCAG 149
      |||||
Qy 121 LeuIleGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAla 140
      |||||
Db 150 CTGATCCAGCTCTACCGAGGATGCCAGCAATGCCATGGTGAGGGCTTCGAGTGGCT 209
      |||||
Qy 141 TyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeu 160
      |||||
Db 210 TATAAGCGGCAGCTGCTGACCATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCTC 269
      |||||
Qy 161 AsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrVal-ProGluLysVa 180
      |||||
Db 270 AATGACCAAGGTGATGAACATGATGGAGACCTGGTCATGGACACAGTCCCTGAAAAGGT 329
      |||||
Qy 180 lHisPheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVal 200
      |||||
Db 330 GCATTCTTCAATAGTTCTTCTATGATAAACTCCGTACCAAGGGTTATGATGGGTGAA 389
      |||||
Qy 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeu-IleProIleHisL 220
      |||||
Db 390 AAGGTGGACCAAAACGTTACATCTTCAATAAGAGGCTACTGCTAAATCCCATCCACC 449
      |||||
Qy 220 euGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheA 240
      |||||
Db 450 TGGAGGTGATGGTGGCTCATCTCTGTGTGATGTGAGGCGACGACCATCCTATTGTG 509
      |||||
Qy 240 sPserGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyr-LeuGlnAla 259
      |||||
Db 510 ACTCGCAGCGTAACCTAAACCGCGCTGCCCTAAGCATATTCCTCAAGTATCTACAGGCA 569
      |||||
Qy 260 GluAlaValLysLysAspArgLeuAspPhe 269
      |||||
Db 570 GAGCGGTAAAGAAAGACCGGATGGATTTC 599
      |||||

RESULT 12
AW604359
LOCUS RC1-CT0249-290100-022-c03 CT0249 Homo sapiens cDNA, mRNA linear EST 23-MAR-2000
DEFINITION AW604359
ACCESSION AW604359
VERSION AW604359.1 GI:7309100
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 620)
JOURNAL HCGP http://www.ludwig.org.br/ORESTES.
COMMENT The FAPESP/LICR Human Cancer Genome Project
Unpublished (1999)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1st2=RC1-CT0249-
290100-022-c03&t3=2000-01-29&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 583.
High quality Location/Qualifiers
1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0249"

FEATURES
source
```

/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 161 a 146 c 163 g 150 t
ORIGIN

Alignment Scores:
Pred. No.: 1,39e-112 Length: 620
Score: 951.00 Matches: 191
Percent Similarity: 91.04% Conservative: 2
Best Local Similarity: 90.09% Mismatches: 2
Query Match: 53.67% Indels: 17
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x AW604359 (1-620)

QY 61 AlaglyGlnHisSerProLeuArgGluGluHisValThrCysValGlnSerIleLeuAsp 80
Db 12 TCAGAGGACATTCG-----GAC 29
QY 81 GluPheLeuGlnThrTyrglySerLeuIleProLeuSerThrAspGluValValGluLys 100
Db 30 GAATTCCTCAACAGTATGCGCCATACCCCTCAGCACTGATGAGGTAGAGAAG 89
QY 101 LeuGluAspIlePheGlnGlnGluPheSerThrProSerArglyGlyLeuValLeuGln 120
Db 90 CTGAGGACATTTCCAGCAGGAGTTTCCACCCCTCCAGAGGGCCCTGGTGTGCAG 149
QY 121 LeftGlnSerTyrglnArgMetProGlyAsnAlaMetValArgGlyPheArgValAla 140
Db 150 CTGATCCAGTCTTACCAGCGGATGCCAGGCAATGCCATGTGTGAGGGCTTCCAGTGGCT 209
QY 141 TyrlysArgHisValLeuThrMetAspAspLeuGlyThrLeuTyrglyGlnAsnTrpLeu 160
Db 210 TATAAGCGGCACGTGCTGACCATGGATGACTTGGGACCTTGTATGCAGAGAAGTGGCTC 269
QY 161 AsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThrVal-ProGluLysVa 180
Db 270 AATGACCATGGTATGACATGTATGGAGACCTGTGTCATGGACACAGTCCCTGAAAGGT 329
QY 180 lHisPheAsnSerPhePheTyrglyAspLysLeuArgThrLysGlyTyrglyVally 200
Db 330 GCATTTCTTCAATAGTTTCTTATGATAAATCCCGTACCAGGGTATGATGGGTGAA 389
QY 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeu-IleProIleHisL 220
Db 390 AAGGTGGACCAAAACGTGTACATCTTCAATAAGGAGCTACTGCTAAATCCCATCCACC 449
QY 220 euGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrglyPheA 240
Db 450 TGGAGGTGCAATGGTCCCTCATCTCTGTTGATGTGAGGCGACGACCATCATTTTGG 509
QY 240 spSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrglyGlnAla 259
Db 510 ACTCGAGCGTAACCTAAACCGCGCTGCCCTAAGCATATTGCCAAGTATCCCTACAGGCA 569
QY 260 GluAlaValLysLysAspArgLeuAspPhe 269
Db 570 GAGCGGTAAGAAAGACCGGATGGATTTC 599

RESULT 13
BE077141
LOCUS
DEFINITION RC5-BT0604-150200-031-E07 BT0604 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE077141
VERSION BE077141.1 GI:8427661
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 572)

AUTHORS

Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., D.H., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

FEATURES

source
1..572
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0604"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 140 a 143 c 149 g 140 t
ORIGIN

Alignment Scores:

Pred. No.: 8,79e-110 Length: 572
Score: 929.00 Matches: 175
Percent Similarity: 88.83% Conservative: 0
Best Local Similarity: 88.83% Mismatches: 0
Query Match: 52.43% Indels: 22
DB: 10 Gaps: 1

US-09-848-852A-3 (1-332) x BE077141 (1-572)

QY 136 GlyPheArgValAlaTyrlysArgHisValLeuThrMetAspLeuGlyThrLeuTy 155
Db 26 GCGTTTCGAGTTGCTTATAAGCGGCACGTCGCACCATGGATGACCTGGGACCTTGTAT 85
QY 156 GlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrglyAspLeuValMetAspThr 175
Db 86 GGACAGAACTGGCTCAATGACGAGGTGATGAACATGATGGAGACCTGTGTCATGGACACA 145
QY 176 ValProGluLysValHisPhePheAsnSerPhePheTyrglyAspLysLeuArgThrLysGly 195
Db 146 GTCCCTGAAAGGTGCATTTCTTCAATAGTTTC----- 178
QY 196 TyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeu 215
Db 179 -----TTCAATAAGGAGCTACTGCTA 199
QY 216 IleProIleHisLeuGluValHisTrpSerLeuIleSerValAspValArgArgThr 235
Db 200 ATCCCCATCCACCTGGAGGTGATGGTCCCTCATCTCTGTGTGTGAGGCGACGCACC 259

```

QY 236 IleThrTyrPheAspSerGlnArgThrLeuAsnArgArgCysProlyshisIleAlaLys 255
|||||
Db 260 ATCACTATTTCGACTCGCAGCGTACCTTAACCGCGCTGCCCTAAGCATATTGCCAAG 319
|||||
QY 256 TyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGly 275
|||||
Db 320 TATCTACAGCAGAGCGGCTAAGAAAGACCGACTGGATTTCACAGCGCTGGAAGGT 379
|||||
QY 276 TyrPheLysMetAsnValAlaArgGlnAsnAspSerAspCysGlyAlaPheValLeu 295
|||||
Db 380 TACTCTCAAATGAATGTGCCAGGAGGAGATAATGACAGTGACTGGTGCTTTGTGTG 439
|||||
QY 296 GlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetPro 315
|||||
Db 440 CAGTACTGCAAGCATCTGCCCTGTCTCAGCCATTTCAGTTCACCCAGCAGGACATGCC 499
|||||
QY 316 LysLeuArgArgGlnIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
|||||
Db 500 AAACCTCGTCGCAGACTCTACAAGGAGCTGTGTCACTGCAAACTCACTGTG 550

RESULT 14
AL578220/c AL578220 626 bp mRNA linear EST 16-FEB-2001
LOCUS AL578220 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK002YD13 3
DEFINITION prime, mRNA sequence.
ACCESSION AL578220
VERSION AL578220.1 GI:12942089
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Pollayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
Location/Qualifiers
1..626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK002YD13"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 142 a 143 c 165 g 164 t 12 others
ORIGIN

Alignment Scores:
Pred. No.: 4..54e-109 Length: 626
Score: 924.00 Matches: 178
Percent Similarity: 90.8% Conservative: 1
Best Local Similarity: 90.36% Mismatches: 18
Query Match: 52.14% Indels: 1
DB: 9 Gaps: 0

US-09-848-852A-3 (1-332) x AL578220 (1-626)

QY 136 GlyPheArgValAlaTyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyr 155
|||||

```

```

Db 625 GCGTTCCGAGTGGCTCATAGCGCGCAGCTGCTGACCATGRATGACATGCGGGACCTTGTAT 566
QY 156 GlyGlnAsnTrpLeuAsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThr 175
|||||
Db 565 GGACAGAACTGGCTCAATGASCAGGTGATGAACATGTATGGAGACCTGCTCATGGACACA 506
|||||
QY 176 ValProGluLysValHisPhePheAsnSerPhePheTyrAspLysLeuArgThrLysGly 195
|||||
Db 505 GTCCCTGAAAGAGTGCATTTCGCAATAGTCTTCTATGATAAACTCCGTACCAAGGT 446
|||||
QY 196 TyrAspGlyValLysArgTrpThrLysAsnValAspIlePheAsnLysLysLeuLeuLeu 215
|||||
Db 445 TATGATGGGBKAAAGGTGGACCAAAACGTCGACATCTCAATAAAGAGCTACTGCTA 386
|||||
QY 216 IleProIleHisLeuGluValHisTrpSerLeuIleSerValAspValArgArgThr 235
|||||
Db 385 ATCCCATCCACTGGAGGTGCTTGTGCTCCATCTCTGTGTGTAGTGAGGCGACGACC 326
|||||
QY 236 IleThrTyrPheAspSerGlnArgThrLeuAsnArgArgCysProlyshisIleAlaLys 255
|||||
Db 325 ACCACCTATTTTCACTCNCACCNATACCTAAACCCGCTGCCCTAAGCATATTGCCAAG 266
|||||
QY 256 TyrLeuGlnAlaGluAlaValLysLysAspArgLeuAspPheHisGlnGlyTrpLysGly 275
|||||
Db 265 TATCTACAGCAGAGCGGCTAAGAAAGACCGACTGGATTTCACCCAGCGCTGGAAGGT 206
|||||
QY 276 TyrPheLysMetAsnValAlaArgGlnAsnAspSerAspCysGlyAlaPheValLeu 295
|||||
Db 205 TACTCTCAAATGAATGTGCCAGGAGGAGTGTGTCACTGCTGCTGCTGCTGCTGCTG 146
|||||
QY 296 GlnTyrCysLysHisLeuAlaLeuSerGlnProPheSerPheThrGlnGlnAspMetPro 315
|||||
Db 145 CAGTACYCGAAGCATCTGCCCTGTCTCAGCCATTTCAGTTCACCCAGCAGGACATGCC 86
|||||
QY 316 LysLeuArgArgGlnIleTyrLysGluLeuCysHisCysLysLeuThrVal 332
|||||
Db 85 AAATT-CGTGCGCAGACTCTACAAGGAGCTGTGTCACTNCAAACTCACTGTG 36
|||||

RESULT 15
BI152305 602917742F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5067995 5',
LOCUS mRNA sequence.
DEFINITION BI152305
ACCESSION BI152305
VERSION BI152305.1 GI:14612306
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1183 row: d column: 12
High quality sequence stop: 523.
Location/Qualifiers
1..870
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5067995"
/tissue_type="spontaneous tumor, metastatic to mammary."
FEATURES
source

```

Stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: PCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 232 a 247 c 190 g 201 t
ORIGIN

Alignment Scores:
Pred. No.: 2.08e-107 Length: 870
Score: 913.00 Matches: 190
Percent Similarity: 88.79% Conservative: 8
Best Local Similarity: 85.20% Mismatches: 13
Query Match: 51.52% Indels: 12
DB: 13 Gaps: 3

US-09-848-852a-3 (1-332) x BII52305 (1-870)

```
Qy 121 LeuileGlnSerTyrGlnArgMetProGlyAsnAlaMetValArgGlyPheArgValAla 140
Db 1 CTGATCCAGTCTTATCAGCGGATGCCAGGCAACGCTATGGTAAGGGGCTTCCGGGTATCC 60
Qy 141 TyrLysArgHisValLeuThrMetAspLeuGlyThrLeuTyrGlyGlnAsnTrpLeu 160
Db 61 TATAAGCGACACGCTGCTCACCATGGATGACTTGGGTACCTTATATGGACAGAACTGGCTC 120
Qy 161 AsnAspGlnValMetAsnMetTyrGlyAspLeuValMetAspThrValProGluLysVal 180
Db 121 AATGACCAGGTGATGAACATGTATGGAGATCTGGTCATGGACACAGTCCCTGAAAGGTG 180
Qy 181 HisPhe-PheAsnSerPhePheTyrAspLysLeuArgThrLysGlyTyrAspGlyVally 200
Db 181 CATATCTTCAACAGCTTCTTCATGATATAACCTCCGTACCAAGGTTATGATGGGGTAAA 240
Qy 200 sArgTrpThrLysAsnValAspIlePheAsnLysGluLeuLeuLeuIleProIleHisLe 220
Db 241 GAGGTGGACCAAAAATGTGGACATCTTCAATAAGGAATTACTGCTAATCCCATCCATCT 300
Qy 220 uGluValHisTrpSerLeuIleSerValAspValArgArgThrIleThrTyrPheAs 240
Db 301 GGAGTGCACTGGTCCCTTATCTCAGTTGATGTAAGGCGCAGTACCACCATCTATTTTGA 360
Qy 240 pSerGlnArgThrLeuAsnArgCysProLysHisIleAlaLysTyrLeuGlnAlaG1 260
Db 361 CTCCACGGAACCTAAATCGCGCTGCCCTAAGCATATTGCCAAGTATCTACACGCCCA 420
Qy 260 uAlaValLysLysAspArgLeuAspPhe-----HisGln-GlyTrp-LysGlyTyrPhe 277
Db 421 GACGGCCAGTCAAAACACGACCGACTGGACTATCCATCAGCGGCTGGACAAGGTACTTC 480
Qy 278 LysMetAsnValAla-ArgGlnAsn---AsnAspSerAspCys-GlyAla---PheValL 295
Db 481 ACATGAATGTGGCCAGCGGACAGATACACTGACAGTACTGCTGGATGCCCTTATGTAT 540
Qy 295 euGlnTyrCys-LysHisLeuAlaLeu-SerGlnPro-PheSerPheThrGlnGlnAspM 314
Db 541 CACAGTACTGCCAAGCACCTGGCCCTGATCTCAGCCACTTCAGCTTCACCCACGAGGACA 600
Qy 314 etProLysLeuArgArgGlnIleTyrLysGluLeuCysHisCysLysLeuThr 331
Db 601 TGCCCAAACTCCGTCGTGATGTACAAGGAAGTGTGTCTACTGCAAACTCACT 653
```

Search completed: December 16, 2002, 23:01:05
Job time : 2428 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2002, 14:18:34 ; Search time 397 Seconds
(without alignments)
11294.027 Million cell updates/sec

Title: US-09-848-852A-4
Perfect score: 1991
Sequence: 1 GCCCTCCCTGTCCCGAGCC.....TGCCTGTGCTAATAAGATC 1991

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT.*		
2:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT.*		
3:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT.*		
4:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT.*		
5:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT.*		
6:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT.*		
7:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT.*		
8:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT.*		
9:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT.*		
10:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT.*		
11:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT.*		
12:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT.*		
13:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT.*		
14:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT.*		
15:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT.*		
16:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT.*		
17:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT.*		
18:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT.*		
19:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT.*		
20:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT.*		
21:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT.*		
22:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT.*		
23:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT.*		
24:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1925.4	96.7	2080	22	AAI58425 Human polynucleoti
2	1843.4	92.6	2206	22	AAI58431 Nucleotide sequenc
3	1813.4	91.1	2029	21	AAI58426 Human polynucleoti
4	1667	83.7	1760	21	AAZ43782 Human fetal brain
5	1568.2	78.8	1804	22	AAI60211 Human polynucleoti
6	1568.2	78.8	1804	22	AAI60212 Human polynucleoti
7	1451.2	72.9	1934	22	AAI05502 Human secreted pro
8	1440.8	72.4	1933	22	ABQ54216 Human ovarian anti
9	1357.4	68.2	1507	22	AAI05541 Human secreted pro

10	567.4	28.5	650	21	AAC59569 Human secreted pro
11	551	27.7	695	22	AAI05540 Human secreted pro
12	486.4	24.4	5318	22	AAF55961 Human eukaryotic i
13	468.4	23.5	870	22	AAI94288 Human neuroblastom
14	442	22.2	503	22	AAI25529 Human ovarian PCR-
15	362.2	18.2	820	22	AAI93958 Human neuroblastom
16	344.6	17.3	714	22	AAH03368 Human CDNA clone (
17	344.6	17.3	1578	22	AAH13708 Human CDNA sequenc
18	291.2	14.6	358	22	AAI84371 Human polynucleoti
19	229.6	11.5	1938	22	AAH99558 Human protein enco
20	208.2	10.5	529	22	AAI05542 Human secreted pro
21	181	9.1	181	21	AAC08625 Human secreted pro
22	177	8.9	208	21	AAA45202 Human secreted exp
23	167.4	8.4	855	22	AAI94287 Human neuroblastom
24	115.8	5.8	501	22	AAI61344 Human immune/haema
25	98.4	4.9	2512	22	AAF54827 Nucleotide sequenc
26	92	4.6	435	24	ABL81757 Human ovarian canc
27	91.8	4.6	666	23	ABL71205 DNA encoding novel
28	86	4.3	833	22	AAI93957 Human neuroblastom
29	84.8	4.3	106	24	ABL81741 Human ovarian canc
30	72	3.6	14551	24	ABL34585 Human metastasis a
31	70.8	3.6	11745	24	ABK28332 DNA transcription
32	70.6	3.5	147	22	AAI61721 Soybean 515002 reg
33	70.6	3.5	150	22	AAI61446 Soybean 240017 reg
34	70.6	3.5	127197	22	AAI61370 Soybean 515002 reg
35	70.6	3.5	335913	22	AAI61371 Soybean 240017 reg
36	70.6	3.5	335913	22	AAI61371 Soybean 240017 reg
37	70.6	3.5	335913	22	AAI61372 Soybean 240017 reg
38	70.6	3.5	335913	22	AAI61372 Soybean 240017 reg
39	70.4	3.5	284	22	AAI61568 Soybean 318013 reg
40	70.4	3.5	513445	22	AAI61373 Soybean 318013 reg
41	70.2	3.5	274	19	AAV16132 Microsatellite mar
42	70	3.5	51952	19	AAV26084 Tomato pest resist
43	69.8	3.5	6298	22	AAF75767 Promoter of gene r
44	69.6	3.5	4049	23	ABL50553 Mial2 nucleotide s
45	69.6	3.5	8952	22	RAA46446 Tumour suppressor

ALIGNMENTS

RESULT 1
AAI58425
ID AAI58425 standard; CDNA; 2080 BP.
XX
AC AAI58425;
XX
DT 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 628.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
PN WO200153312-A1.
XX
PD 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR P-PSDB; AAM39269.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 628; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide of polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 2080 BP; 472 A; 567 C; 570 G; 471 T; 0 other;
XX
XX Query Match 96.7%; Score 1925.4; DB 22; Length 2080;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1967; Conservative 0; Mismatches 1; Indels 21; Gaps 2;
XX
QY 1 GCCTCCCTGTCGCCGACCTCTTTTCATGCTCAGCAAGTGAAGAGGAGGAGGAGG 60
DB 93 GCCTCCCTGTCGCCGACCTCTTTTCATGCTCAGCAAGTGAAGAGGAGGAGGAGG 152
QY 61 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
DB 153 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 212
QY 121 GGAGTCAGCTGGGAACCTCCAGGCGGCCCGCCCTTCCCGCCCGCCACTCATCGAAAACCT 180
DB 213 GGAGTCAGCTGGGAACCTCCAGGCGGCCCGCCCTTCCCGCCCGCCACTCATCGAAAACCT 272
QY 181 GCTCACAGCGCCGCGCGGAGCCATGAGAGCTTCGGAGTCTGCTACTCAAAAAGCA 240
DB 273 GCTCACAGCGCGCGCGGAGCCATGAGAGCTTCGGAGTCTGCTACTCAAAAAGCA 332
QY 241 CCTCGCTGACATTCCTACTGGAAGCTTTGGGGGCGCCACCAGGCGCGCGGCGGCGCTCG 300
DB 333 CCTCGCTGACATTCCTACTGGAAGCTTTGGGGGCGCCACCAGGCGCGCGGCGGCGCTCG 392
QY 301 CACACCCCAAGAACCATCTTTTCAACCCAGCAAGGGGTTGCGAGCCACAGTGCATPCCC 360
DB 393 CACACCCCAAGAACCATCTTTTCAACCCAGCAAGGGGTTGCGAGCCACAGTGCATPCCC 452
QY 361 CCTGTTCTGCTTTTGTAGCTCCCGCGGGGCGCCACTTCCACCCGCGCTGGGTCTGCTAGTG 420
DB 453 CCTGTTCTGCTTTTGTAGCTCCCGCGGGGCGCCACTTCCACCCGCGCTGGGTCTGCTAGTG 512
QY 421 CTCTCATGGCTGAGGATGGGTGAGAGGCTTCCACAGTGCCTCTGCGGCCCGCCCATGG 480
DB 513 CTCTCATGGCTGAGGATGGGTGAGAGGCTTCCACAGTGCCTCTGCGGCCCGCCCATGG 572
QY 481 AGGAAGATGACATGAGTGGACATCCAAAGTCTCTCTGAGACCTTGACTCGGGGCTCTCTTT 540
DB

DB 573 AGGAAGATGACATGAGTGGACTCCAAAGTCTCTCTGAGCCCTGACTCGGGGCTCTCTTT 632
QY 541 CATGTACTCTGCCCAACGGTTTTGGGGGCAATCTCTGGCCAGAGGAGGAGCGCA-CTTGG 599
DB 633 CATGTACTCTGCCCAACGGTTTTGGGGGCAATCTCTGGCCAGAGGAGGAGCGCACTTGG 692
QY 600 CACCCCTGATGCGCAGCATCTCTCATCAGCAATGTGTGAGCATGCGGGGAGCATGTGCCCC 659
DB 693 CACCCCTGATGCGCAGCATCTCTCATCAGCAATGTGTGAGCATGCGGGGAGCATGTGCCCC 752
QY 660 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGGAGGAGGAGGAGGAGG 719
DB 753 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGGAGGAGGAGGAGGAGG 812
QY 720 AAGCGGCCAGCAGACAGCCCTTGGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGG 779
DB 813 AAGCGGCCAGCAGACAGCCCTTGGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGG 872
QY 780 ACGAATTCCTTCAAAACGTATGGCAGCCTCATACCCCTCAGCACTGTAGTAGAGGAGGAG 839
DB 873 ACGAATTCCTTCAAAACGTATGGCAGCCTCATACCCCTCAGCACTGTAGTAGAGGAGGAG 932
QY 840 AGCTGGAGGACATTTTCCAGCAGAGTTCACCCCTTCCAGGAGGAGGAGGAGGAGGAGGAG 899
DB 933 AGCTGGAGGACATTTTCCAGCAGAGTTCACCCCTTCCAGGAGGAGGAGGAGGAGGAGGAG 992
QY 900 AGCTGATCCAGTCTTACACGCGATGCGAGCAATGCGATGCGAGGAGGAGGAGGAGGAGGAG 959
DB 993 AGCTGATCCAGTCTTACACGCGATGCGAGCAATGCGATGCGAGGAGGAGGAGGAGGAGGAG 1052
QY 960 CTTATAGCGCGCACGCTGTCACCATGGATGACTTGGGAGCTTGTATGAGACAGCAACTGGC 1019
DB 1053 CTTATAGCGCGCACGCTGTCACCATGGATGACTTGGGAGCTTGTATGAGACAGCAACTGGC 1112
QY 1020 TCAATGACAGGATGATGAACATGTATGGAGACCTGTCATGGACACAGTCCCTGAAAAGG 1079
DB 1113 TCAATGACAGGATGATGAACATGTATGGAGACCTGTCATGGACACAGTCCCTGAAAAGG 1172
QY 1080 TGCAATTCCTCAATAGTTCTTCTATGATAAAGTCCGTACCAAGGGTTATGATGGGGTGA 1139
DB 1173 TGCAATTCCTCAATAGTTCTTCTATGATAAAGTCCGTACCAAGGGTTATGATGGGGTGA 1232
QY 1140 AAAGGTGGACCAAAACGTCGACATCTTCAATAGGAGCTACTGCTAATCCCATCCACCC 1199
DB 1233 AAAGGTGGACCAAAACGTCGACATCTTCAATAGGAGCTACTGCTAATCCCATCCACCC 1292
QY 1200 TGAGGTGCAATTTGGTCCCTCATCTCTGATGTGAGGCGACGACCATCACTATTTTG 1259
DB 1293 TGGAGGTGCAATTTGGTCCCTCATCTCTGATGTGAGGCGACGACCATCACTATTTTG 1352
QY 1260 ACTCGAGCGCTACCCCTAAACCGCGCTGCGCTTAAGCATATTGCAAGTATCTACAGGCG 1319
DB 1353 ACTCGAGCGCTACCCCTAAACCGCGCTGCGCTTAAGCATATTGCAAGTATCTACAGGCG 1412
QY 1320 AGCGGTAAAGAAAGACCGGACTGGATTTCACCCAGGCGTGGAAAGGTTACTTCAAAATGA 1379
DB 1413 AGCGGTAAAGAAAGACCGGACTGGATTTCACCCAGGCGTGGAAAGGTTACTTCAAAATGA 1472
QY 1380 ATGTGGCCAGGAGATAATATGACAGTGTGCTGCTTTTGTGTCAGTACTGCAAGC 1439
DB 1473 ATGTGGCCAGGAGATAATATGACAGTGTGCTGCTTTTGTGTCAGTACTGCAAGC 1532
QY 1440 ATCTGCGCTCTGTCAGCCATTGAGCTTACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1499
DB 1533 ATCTGCGCTCTGTCAGCCATTGAGCTTACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1592
QY 1500 AGATCTACAGGAGCTGTGTGCTACTGCAAACTCACTGTGTGAGGCTGCTGACCCAGAGCCCC 1559
DB 1593 AGATCTACAGGAGCTGTGTGCTACTGCAAACTCACTGTGTGAGGCTGCTGACCCAGAGCCCC 1652
QY 1560 AAGCCCATAAATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1619
DB 1653 AAGCCCATAAATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1712

QY 1620 CTCCTTGCTCTCCCACTCACTTCCTCTTGGTTTTCATATTTAAATGTTCAATTC 1679
Db 1713 CTCCTTGCTCTCCCACTCACTTCCTCTTGGTTTTCATATTTAAATGTTCAATTC 1772
QY 1680 TGTATTTTTCCTTTCAGAGAACTACTGTTGATTTCTGATGTCAGGGGTGGCTACA 1739
Db 1773 TGTATTTTTCCTTTCAGAGAACTACTGTTGATTTCTGATGTCAGGGGTGGCTACA 1832
QY 1740 GAAAGCCCTCTCTCTCTCTGTTGAGGGAGTGGCCCTGCGCTGGGTGGAGCA 1799
Db 1833 GAAAGCCCTCTCTCTCTCTGTTGAGGGAGTGGCCCTGCGCTGGGTGGAGCA 1892
QY 1800 GTCATCCTCCCTTCCCGTGCAGGGAGCAGGAATCAGTCTCGGGGTGGTGGCGGA 1859
Db 1893 GTCATCCTCCCTTCCCGTGCAGGGAGCAGGAATCAGTCTCGGGGTGGTGGCGGA 1952
QY 1860 CAATAGGATCACTGCTGCGAGATCTTCAAACTTTTATATATATATATATATATAT 1919
Db 1953 CAATAGGATCACTGCTGCGAGATCTTCAAACTTTTATATATATATATATATATAT 1992
QY 1920 AT 1979
Db 1993 AT 2052
QY 1980 TCAATAAAG 1988
Db 2053 TCAATAAAG 2061
RESULT 2
AAF54831
ID AAF54831 standard; DNA; 2206 BP.
XX AAF54831;
AC AAF54831;
XX
DT 15-MAY-2001 (first entry)
XX
DE Nucleotide sequence of a human SENP2 polypeptide.
XX
KW Human; de-sentrinase; SENP; SENP1; SENP2; SENP3; chromosome 12q13.1;
KW sentrin-specific protease; sentrin; anti-proliferative agent;
KW anti-viral agent; herpes simplex-1; cytomegalovirus; sentrinisation;
KW PML; tumour suppressor; acute promyelocytic leukaemia; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 52..1758
FT /*tag= a
FT /product= *SENP3*
PN
PN WO200109292-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-US20884.
XX
PR 31-JUL-1999; 99US-0146774.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Gong L, Yeh ETH;
XX
DR WPI; 2001-182947/18.
DR P-PSDB; AAB31977.
XX
PT New polynucleotide encoding de-sentrinase polypeptides, useful as
PT antiproliferative or antiviral agents -
PS Claim 36; Page 120; 122pp; English.
XX
CC The present sequence encodes a human de-sentrinase (SENP) 3 polypeptide.

CC The specification describes SENP1 and SENP2 polypeptides. The SENP1
CC gene is found in chromosome 12q13.1. SENP1 is found in the nucleus,
CC SENP3 is found in the nucleolus, and SENP1 is found in both locations.
CC SENP polypeptides are sentrin-specific proteases that remove sentrin
CC from some sentrinised peptides, but do not affect proteins modified by
CC ubiquitin or NEDD8. SENP polypeptides are used to identify specific
CC modulators of SENP. These modulators are potential anti-proliferative
CC and anti-viral agents such as herpes simplex-1 or cytomegalovirus.
CC SENP polypeptides are used for studying the role of sentrinisation in
CC the biological function of PML, a tumour suppressor implicated in
CC development of acute promyelocytic leukaemia. Fragments of SENP
CC polynucleotides are used as hybridisation probes and amplification
CC primers for detecting gene expression or preparing mutated sequences,
CC also as antisense sequences for inhibiting SENP expression.
XX
SQ Sequence 2206 BP; 491 A; 612 C; 605 G; 498 T; 0 other;

Query Match 92.6%; Score 1843.4; DB 22; Length 2206;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1925; Conservative 0; Mismatches 31; Indels 33; Gaps 3;

QY 1 GCCTCCCTGTCCTCCCGACCTCTCTTTGATGCTCAGCAAGTGAAGAGGAGGAGAGG 60
Db 230 GCCTCCCTGTCCTCCCGACCTCTTTGATGCTCAGCTAGTAGTGAAGAGGAGGAGAGG 289
QY 61 AGGAGGAGGAGGATGAAGATGAAGAGTGGCAGCTTGGAGGCTGCCCCCAAGAT 120
Db 290 AGGAG-----GAAGATGAGGAGGAGTAGCAGCTGGAGGCTACCCCTAGGT 337
QY 121 GGAGTCACTGGGAACCTCCAGCGGCCCTTCCTCCGCCCTCATCATGAAAACCT 180
Db 338 GGGGCAACTGGGGGCTCCAGCGCTCTCGAGCTCTCCGACCTCTCATGAAAACCT 397
QY 181 GCTACAGCGCGCGCGAGCCATGAGAGCTTCCGGATGCTGCTACTCAAAAAGCA 240
Db 398 GCTACAGCGCGCGCGAGCCATGAGAGCTTCCAGATGCTGCTACTCAAAAAGCA 457
QY 241 CCTCGCTGACATTCACCTGGAAGCTTTGGGGGCGCCACCGGGGCGGGGCGGCTCG 300
Db 458 CCTCGCTGACATTCACCTGGAAGCTTTGGGGGCGCCACCGGGGCGGGGCGGCTCG 517
QY 301 CACACCCCAAGAACCATCTTTACCCCAAGAGGGGTGCGACGCCACAGGTGCCATCC 360
Db 518 CACACCCCAAGAACCATCTTTACCCCAAGAGGGGTGCGACGCCACAGGTGCCATCC 577
QY 361 CCTGTTGCTGTTTACCTCCCGGGGCGCCACCTCCACCCCGGCTGGTCTGCTAGTG 420
Db 578 CCTGTTGCTGTTTACCTCCCGGGGCGCCACCTCCACCCCGGCTGGTCTGCTAGTG 637
QY 421 CTCCTCATGCTGAGGATGGGTGAGAGGCTCTCCACCACTGCTGCGGCCCTCCATGG 480
Db 638 CTCCTCATGCTGAGGATGGGTGAGAGGCTCTCCACCACTGCTGCGGCCCTCCATGG 697
QY 481 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTTGACCTGACTGGGCTCTCTTT 540
Db 698 AGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTTGACCTGACTGGGCTCTCTTT 757
QY 541 CATGTACTCTGCCCAACCGTTTGGGGGACATCTGGGCCAAGAGGGAGGCGCA-CTTGG 599
Db 758 CATGTACTCTGCCCAACCGTTTGGGGGACATCTGGGCCAAGAGGGAGGCGCACTTGG 817
QY 600 CACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGACGATCGGGGACCATGTGGCC 659
Db 818 CACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGACGATCGGGGACCATGTGGCC 877
QY 660 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGGCGCTGGGGAGA 719
Db 878 AGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGAGAGGAGGCGCTGGGGAGA 937
QY 720 AAGCGCGCAGCAGAGCCCTTGGCAGAGGAGCATGTGACCTGCTACAGACATCTTGG 779
Db 938 AAGCGCGCAGCAGAGCCCTTGGCAGAGGAGCATGTGACCTGCTACAGACATCTTGG 997

CC C.N.S disorders.
CC Note: the sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2029 BP; 460 A; 560 C; 553 G; 456 T; 0 other;

Query Match 91.1%; Score 1813.4; DB 22; Length 2029;
Best Local Similarity 96.3%; Pred. No. 0;
Matches 1916; Conservative 0; Mismatches 1; Indels 72; Gaps 3;

QY	1	GCCTCCTCTCCCCGACCCCTCTTTTGATGCTCAGCAAGTGAAGAGGAGAGAGAGG	60
DB	93	GCCTCCTCTCCCCGACCCCTCTTTTGATGCTCAGCAAGTGAAGAGGAGAGAGAGG	152
QY	61	AGGAGGAGGAGTGAAGTGAAGAGGAGGAGTGGACCTGGAGGCTGCCCCCAAGAT	120
DB	153	AGGAGGAGGAGTGAAGTGAAGAGGAGGAGTGGACCTGGAGGCTGCCCCCAAGAT	212
QY	121	GGAGTCAGGTGGGAACCTCCCAAGCGCCCGCCCTTCGCCGCCCACTCATCGAAAAACCT	180
DB	213	GGAGTCAGGTGGGAACCTCCCAAGCGCCCGCCCTTCGCCGCCCACTCATCGAAAAACCT	272
QY	181	GCTCAGCGCGCGCCGCGAGCATCAGAGCCTTCGGATGCTGCTACTCAAAAAGCA	240
DB	273	GCTCAGCGCGCGCGCGAGCATCAGAGCCTTCGGATGCTGCTACTCAAAAAGCA	332
QY	241	CCTCGCTGACATTCACACTGGAAGCTTTGGGGCGCCACCGGGCGCGGGGGGCTCG	300
DB	333	CCTCGCTGACATTCACACTGGAAGCTTTGGGGCGCCACCGGGCGCGGGGGGCTCG	392
QY	301	CACACCCGAAGAACCATCTTTCACCCCAAGAGGGGTGCGAGCGCACAGGTGCCATCCC	360
DB	393	CACACCCGAAGAACCATCTTTCACCCCAAGAGGGGTGCGAGCGCACAGGTGCCATCCC	452
QY	361	CCTGTTGCTGTTTACTCCCCCGGGGGCCACCTCCACCGCGGTGGGTCTGCTAGGTG	420
DB	453	CCTGTTGCTGTTTACTCCCCCGGGGGCCACCTCCACCGCGGTGGGTCTGCTAGGTG	512
QY	421	CTCTCATGCTGAGGATGGGTGAGAGGTCTCCACAGTGCCTCTGGGGCCCCCATGG	480
DB	513	CTCTCATGCTGAGGATGGGTGAGAGGTCTCCACAGTGCCTCTGGGGCCCCCATGG	572
QY	481	AGGAAGTGGACTAGGTGGACTTCCAAAGTCTCCTCTGGACCCCTGACTCGGGCCTCTTT	540
DB	573	AGGAAGTGGACTAGGTGGACTTCCAAAGTCTCCTCTGGACCCCTGACTCGGGCCTCTTT	632
QY	541	CATGTACTCTGCCACAGGTTTGGGGGACATCTGGGCGAGAGGGGAGCGCA-CTTGG	599
DB	633	CATGTACTCTGCCACAGGTTTGGGGGACATCTGGGCGAGAGGGGAGCGCACTTGG	692
QY	600	CACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGCC	659
DB	693	CACCCCTGATGCCAGCATCCTCATCAGCAATGTGTGAGCATCGGGGACCATGTGGCC	752
QY	660	AGGAGCTTTTTCAGGGCTCAGATTGGGTCATGSCAGAAGAGGCGAGAGGCCCTGGGAGA	719
DB	753	AGGAGCTTTTTCAGGGCTCAGATTGGGTCATGSCAGAAGAGGCGAGAGGCCCTGGGAGA	812
QY	720	AGCCGGCCAGACACACCCCTTCGAGAGAGGAGCATGTGACCTGCGTACAGAGCATCTGG	779
DB	813	AGCCGGCCAGACACACCCCTTCGAGAGAGGAGCATGTGACCTGCGTACAGAGCATCTGG	872
QY	780	ACGAATTCCTCAAAGTATGGCAGCTCATACCCCTCAGCAGTATGATGAGGTAGTAGAGA	839
DB	873	ACGAATTCCTCAAAGTATGGCAGCTCATACCCCTCAGCAGTATGATGAGGTAGTAGAGA	932
QY	840	AGCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCAGGAAGGCCCTGGTGTTC	899
DB	933	AGCTGGAGGACATTTTCCAGCAGGAGTTTCCACCCCTTCAGGAAGGCCCTGGTGTTC	992
QY	900	AGCTGATCCAGTCTTACAGCGGATGCCAGGCAATGCCAATGCTGAGGGGCTTCCGAGTGG	959
DB	993	AGCTGATCCAGTCTTACAGCGGATGCCAGGCAATGCCAATGCTGAGGGGCTTCCGAGTGG	1052

QY	960	CTTATAAGCGGACACGCTGCTGACCATGATGACTTGGGGACCTTGTATGACAGAACTGGC	1019
DB	1053	CTTATAAGCGGACACGCTGCTGACCATGATGACTTGGGGACCTTGTATGACAGAACTGGC	1112
QY	1020	TCAATGACAGGTGATGAACATGTATGAGACCTGGTCAATGACACAGTCCCTGAAAGG	1079
DB	1113	TCAATGACAGGTGATGAACATGTATGAGACCTGGTCAATGACACAGTCCCTGAAAGG	1172
QY	1080	TGCATTTCTTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGTTTATGATGGGTGA	1139
DB	1173	TGCATTTCTTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGTTTATGATGGGTGA	1232
QY	1140	AAAGGTGGACCAAAAACGCTGGACATCTTCAATAAGGAGCTACTGCTAAATCCCATCCACC	1199
DB	1233	AAAGGTGGACCAAAAACGCTGGACATCTTCAATAAGGAGCTACTGCTAAATCCCATCCACC	1292
QY	1200	TGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTGAGGAGCGACCATCACTATTTTG	1259
DB	1293	TGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTGAGGAGCGACCATCACTATTTTG	1352
QY	1260	ACTCGCAGCTACCCCTAAACCGCGCTGCCCTAAAGCATATTGCCAAGTATCTACAGGCAG	1319
DB	1353	ACTCGCAGCTACCCCTAAACCGCGCTGCCCTAAAGCATATTGCCAAGTATCTACAGGCAG	1412
QY	1320	AGCGGTAAGAAAGAACCGGACTGGATTCCACAGGGCTGGAAAGGTTTACTTCAAAATGA	1379
DB	1413	AGCGGTAAGAAAGAACCGGACTGGATTTCACAGGGCTGGAAAGGTTTACTTCAAAAT--	1470
QY	1380	ATGTGCCAGGAGCAATAATGACAGTGTGGTGTCTTTGTTGTCAGTACTGCAAGC	1439
DB	1471	-----GTACTGCAAGC	1481
QY	1440	ATCTGCCCTCTCTCAGCCATTACGCTTCCACCCAGCAGCATGCCCCAACTCTCGCGC	1499
DB	1482	ATCTGCCCTCTCTCAGCCATTACGCTTCCACCCAGCAGCATGCCCCAACTCTCGCGC	1541
QY	1500	AGATCTACAAAGAGCTGTCTCAGTGCACAACTCAGTGTGAGCCTCGTACCCAGACCCC	1559
DB	1542	AGATCTACAAAGAGCTGTCTCAGTGCACAACTCAGTGTGAGCCTCGTACCCAGACCCC	1601
QY	1560	AGCCCATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC	1619
DB	1602	AGCCCATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAACTCCAGTTCCTTTC	1661
QY	1620	CTCTCTTCCCTCTCCCACTCAGTTCCTTCTGTTTCTTCAATTTAAATGTTTCAATTTTC	1679
DB	1662	CTCTCTTCCCTCTCCCACTCAGTTCCTTCTGTTTCTTCAATTTAAATGTTTCAATTTTC	1721
QY	1680	TGTATTTTCTTTTGTGAGAAATCTTGTGATTTCTGATGTGAGGGGGTGGCTACA	1739
DB	1722	TGTATTTTCTTTTGTGAGAAATCTTGTGATTTCTGATGTGAGGGGGTGGCTACA	1781
QY	1740	GAAAGCCCTCTTCTTCTCTGTTTGCAGGGAGTGTGGCCCTGTGGCTGGGTGGAGCA	1799
DB	1782	GAAAGCCCTCTTCTTCTCTGTTTGCAGGGAGTGTGGCCCTGTGGCTGGGTGGAGCA	1841
QY	1800	GTCACTCTCCCTCTCCCGTGCAGGAGCAGGAATCAGTGTGGGGTGGGTGGCGGA	1859
DB	1842	GTCACTCTCCCTCTCCCGTGCAGGAGCAGGAATCAGTGTGGGGTGGGTGGCGGA	1901
QY	1860	CAATGAGTCACTTGCCTGCCAGATCTTCAAACTTTTATATATATATATATATATATAT	1919
DB	1902	CAATGAGTCACTTGCCTGCCAGATCTTCAAACTTTTATATATATATATATATATATAT	1941
QY	1920	AT	1979
DB	1942	AT	2001
QY	1980	TCAATAAAG	1988
DB	2002	TCAATAAAG	2010

RESULT 4
AAZ43782
ID AAZ43782 standard; cDNA; 1760 BP.
XX
AC AAZ43782;
XX
DT 10-MAR-2000 (first entry)
XX
DE Human fetal brain cDNA clone vb7_1.
XX
KW Human; secreted protein; treatment; nutritional activity; cytokine;
KW cell proliferation; cell differentiation; hematopoiesis regulation;
KW tissue growth; activin; inhibin; chemotactic; chemokinetic; hemostatic;
KW thrombolytic; anti-inflammatory; invasion suppressor; tumor inhibition;
KW gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO9955721-A1.
XX
PD 04-NOV-1999.
XX
PF 23-APR-1999; 99WO-US08504.
XX
PR 24-APR-1998; 98US-0082904.
PR 11-JUN-1998; 98US-0088994.
PR 12-JUN-1998; 98US-0089278.
PR 02-JUL-1998; 98US-0091847.
PR 24-AUG-1998; 98US-0097639.
PR 22-APR-1999; 99US-0097639.
XX
PA (ALPH-) ALPHAGENE INC.
XX
PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
XX
DR WPI; 2000-052801/04.
DR P-PSDB; AAY50911, AAY50912.
XX
PT New polynucleotides encoding secreted human proteins, derived from
PT human fetal brain, adult skin, adult brain, adult heart, adult thymus
PT and adult aorta cDNA libraries.
XX
PS Claim 20a; Page 223; 282pp; English.
XX
CC This invention describes novel human secreted proteins which are encoded
CC by polynucleotides obtained from fetal brain, adult skin, adult brain,
CC adult heart, adult thymus and adult aorta cDNA libraries. The
CC polynucleotides and proteins are predicted to have biological activities
CC which would make them suitable for treating, preventing or ameliorating
CC medical conditions in humans and animals, although no supporting data
CC is given. Suggested activities include nutritional activity, cytokine
CC and cell proliferation/differentiation activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, hematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, hemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumor
CC invasion suppressor activity, and tumor inhibition activity. The
CC polynucleotides are also stated to be useful for gene therapy.
CC AAZ43777-243808 represent the polynucleotides described in the invention
CC which encode the proteins represented in AAY50905-Y50947.
XX
SQ Sequence 1760 BP; 446 A; 444 C; 450 G; 420 T; 0 other;
XX
Query Match 83.7%; Score 1667; DB 21; Length 1760;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1694; Conservative 0; Mismatches 0; Indels 7; Gaps 2;
QY 289 GGGGGGCGCTCGCACACCCCCAAGAACCATCTTACCCCCAGCAGAGGGGTGCGACGCCAC 348
DB 1 GCGGGGCGCTCGCACACCCCCAAGAACCATCTTACCCCCAGCAGAGGGGTGCGACGCCAC 60
QY 349 AGGTGCATCCCCCTGTGTGCTTTTGTGACTCCCGGGGGGCCACCTCCACCCCGGCTG 408

Db 61 AGGTGCATCCCCCTGTGTGCTTTTGTGCTCCCGGGGGCCACCTCCACCCGGGTGG 120
QY 409 GTCTGCTAGGTGCTCTCATGGCTGAGATGGGGTGTGAGAGGTTCTCCACAGTCCCTCTG 468
Db 121 GTCTGCTAGGTGCTCTCATGGCTGAGATGGGGTGTGAGAGGTTCTCCACAGTCCCTCTG 180
QY 469 GGGCCCCCATGGAGAGATGGACTCAGGTGGGACTCCAAAGTCTCTCTGGAGCCCTGACT 528
Db 181 GGGCCCCCATGGAGAGATGGACTCAGGTGGGACTCCAAAGTCTCTCTGGAGCCCTGACT 240
QY 529 CGGGCCCTCTCTTCATGTACTCTGCCAACGGTTTGGGGACAAATCTGGGCCCAAGAGGGG 588
Db 241 CGGGCCCTCTCTTCATGTACTCTGCCAACGGTTTGGGGACAAATCTGGGCCCAAGAGGGG 300
QY 589 AGGCGA-CTTGGCACCCCTGATGCCAGCATCTCTATCAGCAATGTGTGCAGCATCGGG 647
Db 301 AGGCGAGCTTGGCACCCCTGATGCCAGCATCTCTATCAGCAATGTGTGCAGCATCGGG 360
QY 648 ACCATGTGCCCCAGGAGCTTTTTCAGGGCTCAGATTTTGGGCATGGCAGAGAGGCAGAGA 707
Db 361 ACCATGTGCCCCAGGAGCTTTTTCAGGGCTCAGATTTTGGGCATGGCAGAGAGGCAGAGA 420
QY 708 GGCTTGGGAGAAAGCCGCCAGCACAGCCCCCTGCGAGAGGAGCATGTGACCTGCGTAC 767
Db 421 GGCTTGGGAGAAAGCCGCCAGCACAGCCCCCTGCGAGAGGAGCATGTGACCTGCGTAC 480
QY 768 AGAGCATCTTGGACGAATTCCTTCAAACGTATGGGAGCCTCATACCCCTCAGCAGCTGATG 827
Db 481 AGAGCATCTTGGACGAATTCCTTCAAACGTATGGGAGCCTCATACCCCTCAGCAGCTGATG 540
QY 828 AGGTAGTAGAAGCTGGAGGACATTTTCCAGCAGAGATTTTCCACCCCTTCCAGGAAGG 887
Db 541 AGGTAGTAGAAGCTGGAGGACATTTTCCAGCAGAGATTTTCCACCCCTTCCAGGAAGG 600
QY 888 GCCTGTGTGGCAGCTGATCCAGTCTTTACAGCGGATGCCAGCAATGCCATGGTGAGGG 947
Db 601 GCCTGTGTGGCAGCTGATCCAGTCTTTACAGCGGATGCCAGCAATGCCATGGTGAGGG 660
QY 948 GCTTCCAGTGGCTTATAAGCGCAGCTGCTGACCATGATGACTTTGGGAGCTTTGTATG 1007
Db 661 GCTTCCAGTGGCTTATAAGCGCAGCTGCTGACCATGATGACTTTGGGAGCTTTGTATG 720
QY 1008 GACAGAACTGGCTCAATGACCAAGTGATGAACATGTATGGAGACCTGGTCATGGACACAG 1067
Db 721 GACAGAACTGGCTCAATGACCAAGTGATGAACATGTATGGAGACCTGGTCATGGACACAG 780
QY 1068 TCCTGAAAGGTGCAATTTCTCAATAGTTCCTTCTATGATAAATCCGTACCAAGGGTT 1127
Db 781 TCCTGAAAGGTGCAATTTCTCAATAGTTCCTTCTATGATAAATCCGTACCAAGGGTT 840
QY 1128 ATGATGGGTGAAAGGTGGACCAAAACGTGGACATCTTCAATAAGGAGCTACTCTAA 1187
Db 841 ATGATGGGTGAAAGGTGGACCAAAACGTGGACATCTTCAATAAGGAGCTACTCTAA 900
QY 1188 TCCCCATCCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTGTATGTAGGCGCAGCACCA 1247
Db 901 TCCCCATCCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTGTATGTAGGCGCAGCACCA 960
QY 1248 TCACCTATTTTGACTCGCAGCGTACCCCTAAACGCCGCTGCCCCTAAGCATATTGCCAAGT 1307
Db 961 TCACCTATTTTGACTCGCAGCGTACCCCTAAACGCCGCTGCCCCTAAGCATATTGCCAAGT 1020
QY 1308 ATCTACAGCAGAGCGGTAAAGAAAGCCGACTGATTTTCCACCAAGGCTGGAAGGTT 1367
Db 1021 ATCTACAGCAGAGCGGTAAAGAAAGCCGACTGATTTTCCACCAAGGCTGGAAGGTT 1080
QY 1368 ACTTCAAAATGAATGTGGCAGCAGCAATAATACAGTGTGCTGTGCTTTTGTGTTGC 1427
Db 1081 ACTTCAAAATGAATGTGGCAGCAGCAATAATACAGTGTGCTGTGCTTTTGTGTTGC 1140
QY 1428 AGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCAGCTTACCCAGAGGACATGCCCA 1487

inflammation; neurological disorder; Alzheimer's disease; food additive; angioygenic disorder; kidney disorder; gastrointestinal disorder; allergy; pregnancy-related disorder; endocrine disorder; infection; wound healing; cell culture; chemotaxis; vulnerary; binding partner identification; gene therapy; ss.

Homo sapiens.

Key Location/Qualifiers
CDS 441..713
/*tag= a
/product= "Human secreted protein precursor"
441..521
/*tag= b
mat_peptide 522..710
/*tag= c
/product= "Mature human secreted protein"

WO200134767-A2.

17-MAY-2001.

01-NOV-2000; 2000WO-US30036.

05-NOV-1999; 99US-0163576.

27-JUL-2000; 2000US-0221366.

(HUMA-) HUMAN GENOME SCI INC.

Soppet DR, Komatsoulis G, Shi Y, Olsen HS, Ruben SM;

WPI; 2001-316492/33.

P-PSDB; AAE01682.

Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition - Claim 1; Page 417-418; 540pp; English.

AAD05492-AAD05564 represent cDNAs corresponding to 22 human secreted protein genes, and AAE01672-AAE01743 represent the proteins they encode. AAE01744-AAE01763 represent human secreted protein fragments or variants. The secreted proteins and their genes are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 22 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g. rheumatoid arthritis), inflammation, allergies, autoimmune neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angioygenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

Sequence 1934 BP; 423 A; 542 C; 551 G; 418 T; 0 other;

Query Match 72.9%; Score 1451.2; DB 22; Length 1934;
Best Local Similarity 89.2%; Pred. No. 0;

		Matches 1672; Conservative 0; Mismatches 178; Indels 24; Gaps 9;	
QY	57	GAGGAGGAGGAGGAGATGAAGATGAAGAGGAGGAGAGTGCAGCTGGAGGCTGCCCA	116
DB	12	GAAGAGGAGATGAGGATGAAGAGGAGGAGAGTGCAGCTGGAGGCTGCCCA	71
QY	117	AGATGAGTCACTGGGAAACCTCCACGCGCCGCCCTTCCGCGCCCACTCATCGAAAA	176
DB	72	AGATGGGTGAGTGGGAGCTTCCAGAGCGCTCGTCTCCCGCCCTACTCATCGAAAA	131
QY	177	ACCTGCTCAGCGCGCGCGAGCCATGAGAGCCTTCGGATGCTGCTCTACTCAAAA	236
DB	132	ACCTGCTCAGCGCGCGCGAGCCATGAGAGCCTTCGGATGCTGCTGTACTGAAA	191
QY	237	AGCACCTCGCTGACATTCACCTGGAAGCTTTGGGGCGCCACCGGGCGCGCGCGG	296
DB	192	AGCACCTCGCTGACATTCACCTGGAAGCTTTGGGGCGCCACCGGGCGCGCGGAG-	250
QY	297	CTCGCACACCCCAAGAACCATCTTTTCCACCCCAAGAGGGGTCCGACGCGACAGGTGCCA	356
DB	251	CTCGCACACCCCAAGAACCATCTTTTCCACCCCAAGAGGGGTCCGACGCGAGGTGCCA	310
QY	357	TCCCTCTGTTGCTGTTTGAATTCCTCCCGGGGGGCGGCTCCACCGCGGTGGGTCTGCTA	416
DB	311	TCCCTCTGTTGCTGTTTGAATTCCTCCCGGGGGGCGGCTCCACCGCGGTGGGTCTGCTA	370
QY	417	GGTGTCTCATGCTGAGGATGGGTGAGAGGCTCTCCACAGTGCCTCTGGGGCGGCG	476
DB	371	GGTGTCTCATGCTGAGGATGGGTGAGAGGCTCTCCACAGTGCCTCTGGGGCGGCG	430
QY	477	ATGGAGGAAGATGGACTCAGTGGGACTCCAAAGTCTCTCTTGACCTGACTCGGGCTTC	536
DB	431	ATGGAGGAAGATGGATTACGTTGAGTCCAAAGTCTCTCTTGACCTGACTCGGGCTTC	490
QY	537	CTTTCAATGACTCTGCCCAACGGTTTTGGGGGAGCAATCTGGGCCAGAGGGAGCGCAT	596
DB	491	CTCTCTGTACTCTTCCCAATGGCTTTGGGGGAGCAATCTGGGCCAGAGGGAGCGGT	550
QY	597	-TGGCACCCCTGATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGGACCATGTG	655
DB	551	CTGGCACCCCTGATGCCAGCATCTCATCAGCAATGTGTGAGCATCGGGGACCATGTG	610
QY	656	GCCAGGAGCTTTTTCAGGGCTCAGATTTGGGCTGGGCTGGCAGAGGAGAGGCGCTGG	715
DB	611	GCCAGGAGCTTTTTCAGGGCTCAGATCTGGCAGCGCAGAGAGGAGAGGCGCGCG	670
QY	716	GAGAAAGCGGCGAGCAGAGCGGCTGCGAGAGGAGCATGTGACCTGCGTACAGAGCATC	775
DB	671	GAGAAAGCGGCGAGCAGAGCGGCTGCGGAGGAGCATGTGACCTGCGTACAGAGCATC	730
QY	776	TTGGAGCAATTCCTTCAACGATGGCAGCTCATACCCCTCAGCACTCATGAGGTAGTA	835
DB	731	TTGGAGCAATTCCTTCAACGATGGCAGCTCATACCCCTCAGCACTCATGAGGTAGTA	790
QY	836	GAGAGCTGGAGGAGCATTTTCCAGGAGGATTTTCCACCCCTTCCAGAGGCGCTGGTG	895
DB	791	GAGAAATTAGAGCATTTTCCAGGAGGATTTTCCACCTTCCAGGAGGCGCTGGTG	850
QY	896	TTGAGCTGATCAGTCTTACAGGAGGATGCCAGGCAATGCCATGGTGGAGGCTTCCGA	955
DB	851	CTGAGCTGATCAGTCTTACAGGAGGATGCCAGGCAATGCCATGGTGGAGGCTTCCGG	910
QY	956	GTGGCTTAAAGCGGCGCTGCTGACCATGGATGACTTGGGACCTGTGTATGACAGAAC	1015
DB	911	GTGGCTTAAAGCGGCGCTGCTGACCATGGATGACTTGGGACCTGTGTATGACAGAAC	970
QY	1016	TGGCTCAATGACGAGTGAACATGTATGGAGAGCTGGTGCATGGACAGTGCCTGAA	1075
DB	971	TGGCTCAACGAGGAGTGAACATGTATGGAGAGCTGGTGCATGGACAGTGCCTGAA	1030
QY	1076	AAGGTGCATTTCTTCAATAGTTTCTTATGATAAATCCGTTACCAAGGGTTATGATGG	1135
DB	1031	AAGGTGCATTTCTTCAATAGTTTCTTATGATAAATCCGTTACCAAGGGTTATGATGG	1090

Dbb 72 AGATGGGGTCAGCTGGGAGCTTCCACGGGCTCGTCTCTCCCGCCCTACTCATCGAAAA 131
Qy 177 ACCTGCTCAGAGCCGCGCCGAGCCATGAGAGCCTTCGGATGCTGCTTACTCAAAA 236
Dbb 132 ACCTGCTCCAGCGCGCCGAGCCATGAGAGCCTTCGGATGCTGCTTACTCGAAA 191
Qy 237 AGCACCTCGCTGACATTTCCACTGGAAGCTTTGGGGGGCCACCGGGGCCGCGCGGGC 296
Dbb 192 AGCACCTCGCTGACATTTCCACTGGAAGCTTTGGGGGGCCACCGGGGCCGCGCGGGAG- 250
Qy 297 CTGCGACACCCAGAACCATCTTTACCCACAGCAAGGGGGTGGAGCCACAGGTGCCA 356
Dbb 251 CTGCGACACCCAGAACCATCTTTACCCACAGCAAGGGGGTGGAGCCACAGGTGCCA 310
Qy 357 TCCCCCTGTTGCTGCTTTGACTCCCGCCGGGGCCACTCCACCCGGCTGGGTCTGCTA 416
Dbb 311 TCCCCCTGCTGCTGCTTTGACTCCCGCCGGGGCCACTCCACCCGGCTGGGTCTGCTA 370
Qy 417 GGTGCTCTCATGGCTGAGGATGGGGTGAGAGGCTTCCACAGTGCCTCTGAGCCCTGACTCGGGCTC 536
Dbb 431 GTGGAGGAAGATGAGTTACGGTGAGCTTCCAAAGTCTCCTCTGAGCCCTGACTCTGGCCCTC 490
Qy 537 CTTTCATGTACTCTGCCCAACGGTTTTGGGGGACAATCTGGGCCAGAAAGGGAGCGCACT 596
Dbb 491 CTCTCTGTACTCTTCCCATGAGCTTTGGGGGACCCCGCCGGCCAGNAGGGGAGCGGGT 550
Qy 597 -TGGCACCCCTGATGCCAGCATCTCATCAGCAATGTGCGACATCGGGGACCATGTG 655
Dbb 551 CTGSCACCCCTGATGCCAGCATCTCATCAGCAATGTGCGACATCGGGGACCATGTG 610
Qy 656 GCCCAGAGCTTTTCAGGCGCTCAGATTTGGGCATGCGCAGAGAGGAGAGGCGCTGGG 715
Dbb 611 GCCCAGAGCTATTTTCAGGCGCTCAGATCTGGGCACCCGACAGAGAGGAGCGGCC- 669
Qy 716 GAGAAAGCCGGCAGCAGACCCCTCGAGAGAGGAGGATGTGACCTGCGTACAGAGCATC 775
Dbb 670 GAGNAGCCGGCCAGCAGACCCCTCGGGAGGAGCATGTGACCTGCTGCGAGAGCATC 729
Qy 776 TTGACGAATTCCTTTAAAGCTATGGCAGCCTCATACCCCTCAGCACTGATGAGGTAGTA 835
Dbb 730 TTGATGAATTCCTTTAAAGCTATGGCAGCCTCATACCCCTCAGCAGCGGACGAGGTAGTA 789
Qy 836 GAGNAGCTGAGNAGCATTTTCCAGCAGGAGTTTCCACCCCTTCCAGGAAGGCGCTGGTG 895
Dbb 790 GAGAAATTAGAGGACATTTTCCAGCAGGAGTTTCTTACACCTTCCAGGAAGGCGCTGGTG 849
Qy 896 TTGACGCTGATCAGTCTTACACGCGGATGCCAGGCAATGCCATGTGTGAGGGCTTCCGA 955
Dbb 850 CTGACGCTGATCAGTCTACAGCGGATGCCAGGCAACGCCATGTGTGAGGGCTTCCGG 909
Qy 956 GTGGCTTATAGCGGCGACGTGCTGACCATGGATGACCTGGGCGACCTTATATGACAGAAC 1015
Dbb 910 GTGGCTTACAGCGGCGACGTGCTGACCATGGATGACCTGGGCGACCTTATATGACAGAAC 969
Qy 1016 TGGCTCAATGACAGGTGATGACATGTATGGAGACCTGTGATGACACAGTCCCTGAA 1075
Dbb 970 TGGCTCAACAGCGGTGATGAACATGTACGGAGACCTGTGATGACACAGGTCGCGGAG 1029
Qy 1076 AAGGTGATTTCTCAATAGCTTTCTTATGATAAACTCCGTACCAAGGCTTATGATGG 1135
Dbb 1030 AAGGTGATTTCTCAACAGTTTCTTACAGTAAGCTCGTACCAAGGCTTACGATGGA 1089
Qy 1136 GTGAAAGGTGGACCAAAACGTGGACATCTTCAATAAGGAGTACTGCTTAATCCCATC 1195
Dbb 1090 GTGAAAGGTGGACCAAAACGTGGACATCTTCAATAAGGAGTCTGCTTAATCCCATC 1149
Qy 1196 CACCTGGAGTGTGATGGTCCCTCATCTCTGTTGATGTGAGGCGGACGACCATCATCTAT 1255
Dbb 1150 CACCTGGAGTGTGATGGTCCCTCATCTCTGTTGATGTGAGGCGGCGACCATCATCATGTAT 1209

Qy 1256 TTTGACTCGCAGCGTACCTTAACCCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAG 1315
Dbb 1210 TTTGACTCGCAGCGCACCCTTAACCCGCGCTGCCCTAAGCATATTGCCAAGTATCTACAG 1269
Qy 1316 GCAGAGCGGTAAGAAAGACCGACTGGATTTCCACAGGGCTGGAAAGGTTACTTCAAA 1375
Dbb 1270 GCAGAGCGCATGAAGAAAGACCGCGTGGATTTCCACAGGGCTGGAAAGGTTACTTCAAA 1329
Qy 1376 ATGAATGTGCGCAGGCGAGATAATGACAGTACTGTGGTCTTTTGTGTTGCGAGTACTGC 1435
Dbb 1330 ATGAATGTGCGCAGGCGAGATAATGACAGTACTGTGGTGGCTTTGTGTTACAGTACTGC 1389
Qy 1436 AAGCATCTGCGCCCTGTCTCAGCCATTGAGCTTCAACCCAGCAGGACATGCCCAAACTTCGT 1495
Dbb 1390 AAGCATCTGCGCCCTGTCTCAGCCATTGAGCTTCAACCCAGCAGGACATGCCCAAACTTCGC 1449
Qy 1496 CGCAGATCTACAGAGAGTGTCTCACTGCAAACTCACTGTGTGAGCCTCGTACCCACGA 1555
Dbb 1450 CGCAGATCTACAGAGAGTGTCTCACTGCAAACTCACTGTGTGAGCCTCGTATCCCAGG 1509
Qy 1556 CCCCAAGCCCA----TAAATGGGAAGGAGACATGGAG----TCCCTTCCCAGAAACT 1607
Dbb 1510 CCTCAAGCCCATTTGGTCAATGGGACATGGGGGACCTTCTTCCCAGAAACT 1569
Qy 1608 CAGTTTCTTCTCTCTTCCCTCTTCCCACTCACTTCCCTTTGGTGTTCATATTTAAA 1667
Dbb 1570 CAGTTTCTCTCTCTCTCTCTCC----CCCTCCAGTTCCTCATGTTTTCATATTTAAA 1625
Qy 1668 TGTTCCTCA---ATTCTGATATTTTCTTTTGAGAGAACTTGTGATTTCTGATGT 1723
Dbb 1626 TGTTCCTCAATGATTTCTGATATTTTCTTTTGAGAGAACTTGTGATTTTCTGATGT 1685
Qy 1724 GCAGGGGTGGCTACAGAAAAGCCCTTCTCTCTGTTTGCAGGGAGTGTGGCCCTG 1783
Dbb 1686 TCAGGGGTGGCTACAGAAAGCCCTT--CTCCCTCTGCTAGGAGGAGTGTGGCTGTG 1743
Qy 1784 TGCCCTGGGTGGAGCAGTCACT---CTCCCTTCCCTTCCCTGTCAGGAGCAGGAACTCAGT 1840
Dbb 1744 TGCCCTGGGTGGAGCAGTCACTCCCTTCCCTGTCAGGGGCGGGGAGGAACTCAGT 1803
Qy 1841 GCTGGGGTGGTGGGGGACAAATAGGATCACTGCTGCCAGATCTTCAAACTTTTATATA 1900
Dbb 1804 TGTGGGGTGGTGGGGGGCAA-AGGAGTGTGCTGCCAGATATTTCAAACTTTGTTTA 1862
Qy 1901 TATATATATATATA 1914
Dbb 1863 TACGTATATATAAA 1876
RESULT 9
AAD05541/c
ID AAD05541 standard; cDNA; 1507 BP.
XX
AC AAD05541;
XX
XX
DT 18-JUL-2001 (first entry)
XX
DE Human secreted protein-encoding gene 11 cDNA clone HOF0B27, SEQ ID NO:60.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnery;
KW cell culture; chemotaxis; food additive;
XX binding partner identification; ss.
OS Homo sapiens.


```

XX Key Location/Qualifiers
FH CDS 416..715
FT /*tag= a
FT /product= "Human secreted protein precursor"
FT /note= "Does not include start codon"
FT /partial
FT sig_peptide 416..493
FT /*tag= b
FT mat_peptide 494..712
FT /*tag= c
FT /product= "Mature human secreted protein"
XX
XX WO200134767-A2.
XX
XX 17-MAY-2001.
XX
XX 01-NOV-2000; 2000WO-US30036.
XX
XX 05-NOV-1999; 99US-0163576.
XX 27-JUL-2000; 2000US-0221366.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Soppet DR, Komatsoulis G, Shi Y, Olsen HS, Ruben SM;
XX PI
XX WPI: 2001-316492/33.
XX P-PDSB; AAE01721.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 1: Page 447; 540pp; English.
XX
XX AAD05482-RAD05564 represent cDNAs corresponding to 22 human secreted
XX protein genes, and AAE01672-AAE01743 represent the proteins they encode.
XX AAE01744-AAE01763 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX therapy. Pathological conditions can be diagnosed by determining the
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 22 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of
XX proliferative disorders, cancer, tumours, foetal and developmental
XX abnormalities, haematopoietic disorders, diseases of the immune system,
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX allergies, neurological disorders (e.g., Alzheimer's disease,
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX cardiovascular disorders, angiotonic disorders, kidney disorders,
XX gastrointestinal disorders, pregnancy-related disorders, endocrine
XX disorders, and infections. The proteins can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues, to identify their
XX cognate ligands or binding partners, and in chemotaxis, and can be used
XX as a food additive or preservative to modify storage properties.
XX Antibodies specific for a protein of the invention can be used in
XX alleviating symptoms associated with the disorders mentioned above, and
XX in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX immunosorbent assay (ELISA). The present sequence represents a human
XX secreted protein-encoding cDNA of the invention.
XX
XX Sequence 1507 BP; 357 A; 385 C; 354 G; 401 T; 10 other;
XX
XX Query Match 68.2%; Score 1357.4; DB 22; Length 1507;
XX Best Local Similarity 98.5%; Pred No. 2e-308;
XX Matches 1358; Conservative 9; Mismatches 11; Indels 0; Gaps 0;
XX
XX 558 GGTGTTGGGGGACAATCTGGCCAGAGGGGAGCGCACTTGGCACCCCTGATGCCAGCA 617
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 1467 GGGTCAGAGGAGACTTTGGCCAGAGGGGAGCGCANTTGGCACCCCTGATGCCAGCA 1408

```

QY 1698 AGAATACTTGTTGATTTCTGATGTCAGGGGGTGGCTACAGAAAGCCCTTCTTCC 1757
Db AAAAAAATACCTTGTTGATTTCTGATGTCAGGGGGTGGCTACAGAAAGCCCTTCTTCC 268
QY 1758 TCTGTTTGCAGGGAGTGTGGCCCTGTGGCTGGGTGGAGCAGTCATCCTCCCTTCC 1817
Db TCTGTTTGCAGGGAGTGTGGCCCTGTGGCTGGGTGGAGCAGTCATCCTCCCTTCC 208
QY 1818 CGTCAGGAGCAGGAATCAGTCTGGGGTGGTGGCGGACAAATAGGATCACTGCCTG 1877
Db 207 CGTGCAGGAGCAGGAATCAGTCTGGGGTGGTGGCGGACAAATAGGATCACTGCCTG 148
QY 1878 CCAGATCTTCAAACTTTTATATATATATATATATATATATATATATATATATAT 1935
Db 147 CCAGATCTTCAAACTTTTATATATATATATATATATATATATATATATATATAT 90

RESULT 10
AAC59569
ID AAC59569 standard; cDNA; 650 BP.
XX AC AAC59569;
XX 26-JAN-2001 (first entry)
XX Human secreted protein gene 4 SEQ ID NO:14.
DE Human; secreted protein; diagnosis; neuroprotective; cytotstatic;
XX cardioactive; immunomodulatory; muscular active general; vulnery;
KW gastrointestinal; nephrotropic; antinefactive; gynaecological; and
KW antibacterial; gene therapy; detection; cancer; chromosome marker;
KW chromosome identification; neural disorder; immune disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; wound healing; infectious disease; preservative;
KW food additive; ss.
XX Homo sapiens.
XX OS
XX PN WO200056883-A1.
XX 28-SEP-2000.
PD 16-MAR-2000; 2000WO-US06822.
PF 23-MAR-1999; 99US-0126054.
PR 10-DEC-1999; 99US-0169916.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Komatsoulis G;
PI WPI: 2000-587666/55.
DR P-PSDB; AAB34302.
XX Human secreted proteins and gene sequences encoding them, useful for
PT detecting, preventing, and treating disorders such as cancer,
PT neurological disorders and immune system disorders -
XX Claim 1; Page 340; 429pp; English.
XX The polynucleotide sequences given in AAC59566 to AAC59614 encode the
CC human secreted proteins given in AAB34299 to AAB34347. AAB34348 to
CC AAB34437 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Example of activities include:
CC neuroprotective; cytotstatic; cardioactive; immunomodulatory; muscular
CC active general; gynaecological; and antibacterial. The polynucleotides
CC antinefactive; gynaecological; and antibacterial. The polynucleotides
CC can be used for the detection of various disorders such as cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The secreted proteins can be used to

CC treat disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wound healing, and infectious diseases. The proteins can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. AAC59557 to AAC59565 and AAB34298 represent
CC sequences used in the exemplification of the present invention.
XX SQ Sequence 650 BP; 164 A; 155 C; 149 G; 182 T; 0 other;
Query Match 28.5%; Score 567.4; DB 21; Length 650;
Best Local Similarity 99.7%; Pred. No. 3.1e-123;
Matches 579; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1347 TCACCCAGGCTGGAAGGTTACTTCAAAATGAATGTGGCAGGACAGTAATGACAGTG 1406
Db 10 TCACCCAGGCTGGAAGGTTACTTCAAAATGAATGTGGCAGGACAGTAATGACAGTG 69
QY 1407 ACTGTGTGCTTTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1466
Db 70 ACTGTGTGCTTTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 129
QY 1467 TCACCCAGGAGGACATGCCCCAACTTCTGCGGAGATCTACAGGAGCTGTGCTACTGCA 1526
Db 130 TCACCCAGGAGGACATGCCCCAACTTCTGCGGAGATCTACAGGAGCTGTGCTACTGCA 189
QY 1527 AACTCACTGTGTGAGCTCGTACCCAGACCCCAAGCCATTAATGGAGGAGAGACATG 1586
Db 190 AACTCACTGTGTGAGCTCGTACCCAGACCCCAAGCCATTAATGGAGGAGAGACATG 249
QY 1587 GGAGTCCCTTCCCAAGAACTCCAGTTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1646
Db 250 GGAGTCCCTTCCCAAGAACTCCAGTTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 309
QY 1647 CTTTGTGTTTTCATATTTAAATGTTTCAATTTCTGTAATTTTTTTTCTTTTGAGAGATAC 1706
Db 310 CTTTGTGTTTTCATATTTAAATGTTTCAATTTCTGTAATTTTTTTTCTTTTGAGAGATAC 369
QY 1707 TTGTTGATTTCTGATGTGAGGGGGTGGCTACAGAAAAGCCCTTCTTCTCTGTTTTC 1766
Db 370 TTGTTGATTTCTGATGTGAGGGGGTGGCTACAGAAAAGCCCTTCTTCTCTGTTTTC 429
QY 1767 AGGGAGTGTGGCCCTGTGGCTGGGTGGAGCAGTCATCTCCCTTCCCTGTCAGGG 1826
Db 430 AGGGAGTGTGGCCCTGTGGCTGGGTGGAGCAGTCATCTCCCTTCCCTGTCAGGG 489
QY 1827 AGCAGGAATCAGTGTGGGGTGGGTGGGACAAATAGGATCAGTCGCTGCCAGATCTT 1886
Db 490 AGCAGGAATCAGTGTGGGGTGGGTGGGACAAATAGGATCAGTCGCTGCCAGATCTT 549
QY 1887 CAAACTTTTAT 1927
Db 550 CAAAC-TTTAT 589

RESULT 11
AAB05540
ID AAB05540 standard; cDNA; 695 BP.
XX AC AAB05540;
XX 18-JUL-2001 (first entry)
XX Human secreted protein-encoding gene 11 cDNA clone HOF0B27, SEQ ID NO:59.
DE Human; secreted protein; proliferative disorder; cancer; tumour;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnery;

PA	(GLAX) GLAXO GROUP LTD.	
XX	Greaves DR, Thomsen L, Catchpole IR, Ford MJ;	
XX	WPI; 2001-138151/14.	
XX	New DNA constructs, useful in gene therapy for treating or preventing	
PT	chronic inflammatory reactions such as multiple sclerosis or asthma, as	
PT	DNA vaccines, or in the commercial production of proteins -	
XX	Claim 19; Page 58-62; 62pp; English.	
XX	The present invention describes a DNA construct comprising a	
CC	transcriptional regulatory sequence linked to the gene of interest where	
CC	the regulatory sequence is the human eukaryotic initiation factor 4A1	
CC	(eIF4A1) gene promoter. This can be used in the production of DNA	
CC	vaccines and gene therapy, both of which can be used in the prevention	
CC	and treatment of diseases such as cystic fibrosis, Parkinson's disease,	
CC	cancer, hepatitis B and C, HIV, tuberculosis, HPV, HSV, multiple	
CC	sclerosis, Alzheimer's disease and asthma. It can also be used in	
CC	vaccines for contraception or drug addition.	
XX	Sequence 5318 BP; 1327 A; 1210 C; 1423 G; 1350 T; 8 other;	
XX	Query Match 24.4%; Score 486.4; DB 22; Length 5318;	
XX	Best Local Similarity 94.9%; Pred. No. 7.6e-104;	
XX	Matches 544; Conservative 1; Mismatches 17; Indels 11; Gaps 4;	
QY	1427 CAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCAGCTTACCCAGCAGGACATGCC 1486	
DB		
DB	3827 CAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTCAGCTTACCCAGCAGGACATGCC 3886	
QY	1487 AAATCTCGTGGCAGATCTACAAGGAGCTGTGTCACATGCAAACTCACTGTGTGAGCTCG 1546	
DB		
DB	3887 AAATCTCGTGGCAGATCTACAAGGAGCTGTGTCACATGCAAACTCACTGTGTGAGCTCG 3946	
QY	1547 TACCCAGACCCCAAGCCCAATAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAAC 1606	
DB		
DB	3947 TACCCAGACCCCAAGCCCAATAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAAC 4006	
QY	1607 TCAGTTTCCCTTCCCTCTCTGCTCTTCCACATCCTTCCCTTGGTTTTCATATTAA 1666	
DB		
DB	4007 TCAGTTTCCCTTCCCTCTCTGCTCTTCCACATCCTTCCCTTGGTTTTCATATTAA 4066	
QY	1667 ATGTTTCAATTTTC-TGTAATTTTTTTTCTTTTGAGAGAACTACTTGTGATTTCTGATGTGC 1725	
DB		
DB	4067 ATGTTTCAATTTCTGTAATTTTTTTTCTTTTGAGAGAACTACTTGTGATTTCTGATGTGC 4126	
QY	1726 AGGGGTGGGTACAGAAAAGCCCTTTCTTCCCTCTGTTTGCAGGGGAGTGTGGCCCTGTG 1785	
DB		
DB	4127 AGGGGTGGGTACAGAAAAGCCCTTTCTTCCCTCTGTTTGCAGGGGAGTGTGGCCCTGTG 4186	
QY	1786 G-CCTGGGTGAGAGTCACTTCCCTTCCCTTCCCTGCGGAGCAGGAAATCAGTGCT- 1843	
DB		
DB	4187 GCGGTGGGTGAGAGTCACTTCCCTTCCCTTCCCTGCGGAGCAGGAAATCAGTGATG 4246	
QY	1844 GGGGTGGGTGGGCGCAATAGATCACTCCCTGCCAGATCTTCAACTTTTATATATAT 1903	
DB		
DB	4247 GGGGTGGGTGGGCGCAATAGATCACTCCCTGCCAGATCTTCAACTTTTATATATAT 4306	
QY	1904 AT 1957	
DB		
DB	4307 AT 4366	
QY	1958 --AAATGCCACGGTCTCTGGTCAATAAAG 1988	
DB		
DB	4367 AAAAATGCCACGGTCTCTGGTCAATAAAG 4399	
XX	RESULT 13	
XX	AA194288/c	
XX	ID AA194288 standard; cDNA; 870 BP.	
XX		

QY 936 CCATGCTGAGGGCTTCCGAGTGGCTTATAAGCGGCACCTGCTGACCATGGATGACTTGG 995
Db 451 CCATGCTGAGGGCTTCCGAGTGGCTTATAAGCGGCACCTGCTGACCATGGATGACTTGG 392
QY 996 GGACCTTGATGACAGAACTGGCTCAATGACCGAGTGTATGACATGTATGAGACCTGG 1055
Db 391 GGACCTTGATGACAGAACTGGCTCAATGACCGAGTGTATGACATGTATGAGACCTGG 332
QY 1056 TCATGGACACAGTCCCTGAAAAGGT 1080
Db 331 TCATGGACACAGTCCCTGAAAAGGT 307
RESULT 14
ID AAS25529/c
XX AAS25529 standard; cDNA; 503 BP.
AC AAS25529;
DT 07-NOV-2001 (first entry)
DE Human ovarian PCR-subtracted cDNA library clone #1614.
XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;
KW gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;
KW primer; probe.
OS Homo sapiens.
XX WO200157207-A2.
PN 09-AUG-2001.
PD 05-FEB-2001; 2001WO-US03733.
XX 04-FEB-2000; 2000US-0180403.
PR 28-MAR-2000; 2000US-0192745.
XX (CORI-) CORIXA CORP.
XX Algate PA, Mannion J;
PI WPI; 2001-488879/53.
XX New polynucleotides encoding ovarian tumour proteins, useful for
PT treating ovarian cancer, and as probes, primers, and markers of cancer
PT progression -
XX Example 1; page 375; 378pp; English.
XX The invention comprises compositions used for the therapy and diagnosis
CC of ovarian cancer. The compositions comprise one or more ovarian tumour
CC proteins, their associated polynucleotides, or immunogenic portions of
CC the proteins. The ovarian tumour polynucleotides and polypeptides are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein. They are also useful for inhibiting the development of cancer in
CC a patient with an ovarian tumour DNA or protein by incubating isolated
CC T-cells allowing them to proliferate, and administering to the patient.
CC The sequences can be used as markers for cancer, for example, to monitor
CC ovarian cancer progression. Probes and primers are useful in nucleic acid
CC hybridisation, in detecting the presence of complementary sequences in a
CC given sample, for preparing mutant species and for preparing other
CC genetic constructions. Sequences AAS23820-AAS25231 and AAS25528-AAS25549
CC represent human ovarian tumour protein cDNA clones.
SQ Sequence 503 BP; 120 A; 110 C; 147 G; 126 T; 0 other;
XX Query Match 22.2%; Score 442; DB 22; Length 503;
XX Best Local Similarity 90.8%; Pred. No. 7.3e-94;
XX Matches 503; Conservative 0; Mismatches 0; Indels 51; Gaps 1;
QY 1111 ACTCCGTACCAAGGGTTATGATGGGTGAAAAGGTGGACCAAAACGGTGGACATCTTCAA 1170
|||||

Db 503 ACTCCGTACCAAGGGTTATGATGGGTGAAAAGGTGGACCAAAACGGTGGACATCTTCAA 444
QY 1171 TAAGGAGCTACTGCTAATCCCATCCACCTGGAGGTGCATTGGTCCCTCATCTCTGTGA 1230
Db 443 TAAGGAGCTACTGCTAATCCCATCCACCTGGAGGTGCATTGGTCCCTCATCTCTGTGA 384
QY 1231 TGTGAGGCGACGACCATCACCCTATTTTGTGCTGCGAGCTAGCCTTAAACCGCGGTGCC 1290
Db 383 TGTGAGGCGACGACCATCACCCTATTTTGTGCTGCGAGCTAGCCTTAAACCGCGGTGCC 324
QY 1291 TAAGCATATTTGCCAAGTATCTACAGGCAGAGCGGTAAAGAAAGACCGACTGGATTCCA 1350
Db 323 TAAGCATATTTGCCAAGTATCTACAGGCAGAGCGGTAAAGAAAGACCGACTGGATTCCA 264
QY 1351 CCAGGCTGAAAAGGTTACTTCAAAATGAATGTGGCCAGGCAGATAATGACAGTCACTG 1410
Db 263 CCAGGCTGAAAAGGTTACTTCAAAAT-----GTACTGCAAGCATCTGGCCCTTCA 237
QY 1411 TGGTGCTTTTGTGTTGTCAGTACTGCAAGCATCTGGCCCTTCTTCAGCCATTCACTGTCAC 1470
Db 236 -----GTACTGCAAGCATCTGGCCCTTCTTCAGCCATTCACTGTCAC 195
QY 1471 CCAGCAGGACATGCCCAAACTTCGTCGGCAGATCTACAAGGAGCTGTGCTGCTGCAAACT 1530
Db 194 CCAGCAGGACATGCCCAAACTTCGTCGGCAGATCTACAAGGAGCTGTGCTGCTGCAAACT 135
QY 1531 CACTGTGTGAGCTCGTACCCAGAGCCCAAGCCCAATTAATGGAGGAGACATGGGAG 1590
Db 134 CACTGTGTGAGCTCGTACCCAGAGCCCAAGCCCAATTAATGGAGGAGACATGGGAG 75
QY 1591 TCCCTTCCCAAGAACTCCAGTTCCTTCTTTCCTTCTTTCCTTCTTTCCTTCTTTCCTTCTT 1650
Db 74 TCCCTTCCCAAGAACTCCAGTTCCTTCTTTCCTTCTTTCCTTCTTTCCTTCTTTCCTTCTT 15
QY 1651 GGTTCCTCATATTT 1664
Db 14 GGTTCCTCATATTT 1
RESULT 15
AAI93958/c
ID AAI93958 standard; cDNA; 820 BP.
XX AAI93958;
AC AAI93958;
DT 13-NOV-2001 (first entry)
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 33.
XX Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX Homo sapiens.
OS WO200166719-A1.
PN 13-SEP-2001.
PD 02-MAR-2001; 2001WO-JP01629.
XX 07-MAR-2000; 2000JP-0159195.
PR (CHIB-) CHIBA PREFECTURE.
XX (HISM) HISAMITSU PHARM CO LTD.
PI Nakagawara A;
XX WPI; 2001-565584/63.
XX Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents -
XX

PS Claim 1; Page 59; 2979pp; Japanese.

XX The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.

XX
SQ
Sequence 820 BP; 224 A; 223 C; 170 G; 191 T; 12 other;
XX

Query Match 18.2%; Score 362.2; DB 22; Length 820;

Query Match 10.2%; Score 302.2; DB 2
Best Local Similarity 85.1%; Pred. No. 4.7e-75;

Sequence	Matches	Mismatches	Indels	Gaps
Sequence 1	481	0	70	14
Sequence 2	481	0	70	14
Sequence 3	481	0	70	14
Sequence 4	481	0	70	14
Sequence 5	481	0	70	14
Sequence 6	481	0	70	14
Sequence 7	481	0	70	14
Sequence 8	481	0	70	14
Sequence 9	481	0	70	14
Sequence 10	481	0	70	14
Sequence 11	481	0	70	14
Sequence 12	481	0	70	14
Sequence 13	481	0	70	14
Sequence 14	481	0	70	14
Sequence 15	481	0	70	14
Sequence 16	481	0	70	14
Sequence 17	481	0	70	14
Sequence 18	481	0	70	14
Sequence 19	481	0	70	14
Sequence 20	481	0	70	14
Sequence 21	481	0	70	14
Sequence 22	481	0	70	14
Sequence 23	481	0	70	14
Sequence 24	481	0	70	14
Sequence 25	481	0	70	14
Sequence 26	481	0	70	14
Sequence 27	481	0	70	14
Sequence 28	481	0	70	14
Sequence 29	481	0	70	14
Sequence 30	481	0	70	14
Sequence 31	481	0	70	14
Sequence 32	481	0	70	14
Sequence 33	481	0	70	14
Sequence 34	481	0	70	14
Sequence 35	481	0	70	14
Sequence 36	481	0	70	14
Sequence 37	481	0	70	14
Sequence 38	481	0	70	14
Sequence 39	481	0	70	14
Sequence 40	481	0	70	14
Sequence 41	481	0	70	14
Sequence 42	481	0	70	14
Sequence 43	481	0	70	14
Sequence 44	481	0	70	14
Sequence 45	481	0	70	14
Sequence 46	481	0	70	14
Sequence 47	481	0	70	14
Sequence 48	481	0	70	14
Sequence 49	481	0	70	14
Sequence 50	481	0	70	14
Sequence 51	481	0	70	14
Sequence 52	481	0	70	14
Sequence 53	481	0	70	14
Sequence 54	481	0	70	14
Sequence 55	481	0	70	14
Sequence 56	481	0	70	14
Sequence 57	481	0	70	14
Sequence 58	481	0	70	14
Sequence 59	481	0	70	14
Sequence 60	481	0	70	14
Sequence 61	481	0	70	14
Sequence 62	481	0	70	14
Sequence 63	481	0	70	14
Sequence 64	481	0	70	14
Sequence 65	481	0	70	14
Sequence 66	481	0	70	14
Sequence 67	481	0	70	14
Sequence 68	481	0	70	14
Sequence 69	481	0	70	14
Sequence 70	481	0	70	14
Sequence 71	481	0	70	14
Sequence 72	481	0	70	14
Sequence 73	481	0	70	14
Sequence 74	481	0	70	14
Sequence 75	481	0	70	14
Sequence 76	481	0	70	14
Sequence 77	481	0	70	14
Sequence 78	481	0	70	14
Sequence 79	481	0	70	14
Sequence 80	481	0	70	14
Sequence 81	481	0	70	14
Sequence 82	481	0	70	14
Sequence 83	481	0		

QY 517 TGGACCCCTGACTCGGGCCCTCCTTTCATGTACTCTGCCCAACGGTTTGGGGGACAAATCTG 576

Db 820 TTGGACCTTGATGGGGCCTCCNTTTNATGNATTTGCCAAACGTTTTGGGGG ---AAATTT 764

QY 577 GCCAGAGGGAGCGCACTTGGCACCCCTGATGCCAGCATCCTCATCAGCAATGTGTG 636

Db 763 GGCCCANAGGGGAGCGCAGTTTGCACCCCTTGATCCA--GATCCTTATCAGCAATNTT- 707

QV 637 CAGCATCGGGGACCATGTGGCCCAAGGAGCTTTTTCAGGGCTCAGATTTGGGCATGGCAGA 696

Cf

CGT CAGGATCGGGGACA ---TTTGCCCAAGGAGNTTTTCAGGSNVCAGATTGGCATGCAGAGS 650

D_b

OV 697 AGAGGCAGAGAGGCCCTGGGGAGAAAGCCGGCCAGCACAGCCCCCTGCGAGAGGAGCATGT 756

[illegible]

OV 757 GACCTGGGTACAGAGCATCTTGGACGAA TTCCCTTC-AAA CGTATGGCAGCCTCATACCC 815

QY	757	GACCTGCGGACAGAGGCATCTTGGACGAATTCCTTC-AAAACGATATGACAGCCCTCATACCCC
Db	593	GACCTTGGCGACAGAGCAGCANTTTGGAGCAGCANTTCTTCAAAACGTATGACAGCCCTCATACCCCT

[illegible]

QY	816	TCAGACCTGATGAGGTAGTACAGAACTCGAGGACATTTTCCAGCAGGAGTTTCCACCC	879
D6	533	TCAGACTGATGAGCTACTGACAGACAGACGAGCTTCCAGGACACATTTTTTCCACCC	474

[illegible][illegible][illegible]

936 CCATGGTGAGGGGCTTCCGAGTGGCTTATTAAGCGGCACGTGCTGACCATGGATGACTGG 995

[illegible]

996 GGACCTTGTATGGACAGAACTGGCTCAATGACCAGGTGATGAACATGTATGGAGACCTGG 105

[illegible]

QY 1056 TCATGGACACAGTCCCTGAAAAGGT 108

Search completed: December 16, 2002, 15:38:59

```
search completed: Done
Job time : 418 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2002, 13:18:49 ; Search time 64 Seconds
(without alignments)
9540.522 Million cell updates/sec

Title: US-09-848-852A-4
Perfect score: 1991
Sequence: 1 GCCCTCCTGTCCCGGCC.....TGCCTGTGCTCAATAAGATC 1991

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2.6/ptodata/2/ina/PTOTUS_COMB.seq: *
6: /cgn2.6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70	3.5	51952	3	US-08-947-823-1
2	67.8	3.4	1890	6	Sequence 1, Appli Patent No. 5312912
3	67.2	3.4	51952	3	Sequence 1, Appli Patent No. 5312912
4	67	3.4	480	1	Sequence 4, Appli US-08-282-581-4
5	67	3.4	480	1	Sequence 4, Appli US-08-550-544-4
6	66.2	3.3	1890	6	Sequence 4, Appli Patent No. 5312912
7	65.8	3.3	480	1	Sequence 4, Appli US-08-282-581-4
8	65.8	3.3	480	1	Sequence 4, Appli US-08-550-544-4
9	65.6	3.3	297	3	Sequence 17, Appli US-08-616-368A-17
10	65.6	3.3	297	3	Sequence 17, Appli US-09-054-298-17
11	65.6	3.3	297	4	Sequence 17, Appli US-08-818-655-17
12	65	3.3	17656	4	Sequence 3, Appli US-09-433-579-3
13	64.6	3.2	128	1	Sequence 8, Appli US-07-922-723A-8
14	64.6	3.2	128	1	Sequence 8, Appli US-08-550-544-4
15	64.6	3.2	128	1	Sequence 8, Appli US-08-616-368A-17
16	64.6	3.2	128	1	Sequence 8, Appli US-08-074-275-8
17	64.6	3.2	128	2	Sequence 8, Appli US-07-952-277A-8
18	64.6	3.2	297	1	Sequence 17, Appli US-08-616-368A-17
19	64.6	3.2	297	3	Sequence 17, Appli US-09-054-298-17
20	64.6	3.2	297	4	Sequence 17, Appli US-08-818-655-17
21	64.6	3.2	4254	2	Sequence 7, Appli US-08-443-639-7
22	63.8	3.2	1368	3	Sequence 5, Appli US-08-874-563-5
23	63.8	3.2	1368	3	Sequence 14, Appli US-08-577-483-14
24	63.8	3.2	1253	3	Sequence 7, Appli US-08-577-483-7
25	63.6	3.2	1368	3	Sequence 5, Appli US-08-874-563-5
26	63.6	3.2	1368	3	Sequence 14, Appli US-08-577-483-14
27	63.6	3.2	4253	3	Sequence 7, Appli US-08-577-483-7

28	63.4	3.2	128	1	US-07-922-723A-8	Sequence 8, Appli
29	63.4	3.2	128	1	US-07-799-828C-8	Sequence 8, Appli
30	63.4	3.2	128	1	US-08-074-275-8	Sequence 8, Appli
31	63.4	3.2	128	1	US-08-480-366-8	Sequence 8, Appli
32	63.4	3.2	128	2	US-07-952-277A-8	Sequence 8, Appli
33	62.8	3.2	87350	3	US-08-781-891-79	Sequence 79, Appli
34	62.8	3.2	87543	4	US-09-791-211-3	Sequence 3, Appli
35	62.2	3.1	4254	2	US-08-443-639-7	Sequence 7, Appli
36	62	3.1	15062	1	US-08-207-904-18	Sequence 18, Appli
37	61.4	3.1	15062	4	US-09-004-838-89	Sequence 89, Appli
38	61.4	3.1	87350	3	US-08-781-891-79	Sequence 79, Appli
39	61.4	3.1	87543	4	US-09-791-211-3	Sequence 3, Appli
40	61.4	3.1	246240	2	US-08-724-394A-20	Sequence 20, Appli
41	61.4	3.1	246240	2	US-08-724-394A-21	Sequence 21, Appli
42	61.4	3.1	246240	2	US-08-724-394A-22	Sequence 22, Appli
43	61.2	3.1	3072	4	US-09-522-217-55	Sequence 55, Appli
44	61.2	3.1	15062	4	US-09-004-838-89	Sequence 89, Appli
45	61	3.1	1906	1	US-08-207-904-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1
US-08-947-823-1
; Sequence 1, Application US/08947823
; Patent No. 6114605
; GENERAL INFORMATION:
; APPLICANT: Williamson, Valerie M.
; APPLICANT: Kaloshian, Isghouhi
; APPLICANT: Yaghoobi, Jafar
; APPLICANT: Bodeau, John
; APPLICANT: Milligan, Stephen
; TITLE OF INVENTION: Procedures and Materials for Conferring
; TITLE OF INVENTION: Pest Resistance in Plants
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/947,823
; FILING DATE: 09-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/18802
; FILING DATE: 09-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,191
; FILING DATE: 10-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-070210US
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-947-823-1

Query Match 3.4%; Score 67; DB 1; Length 480;
Best Local Similarity 79.8%; Pred. No. 3.9e-08;
Matches 79; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

[illegible]

QY 1992 ATATATAAAGGACGGGCCCTCGTCGTGCAGATTAAAGAT 1990
||||| | | | | | | | |
Db 245 ATATATATAATGCAGAAGTATCTTTGTAGATTATAAT 207

RESULT 6
5312912-3/c
: Patent No. 5312912

Query Match	3.3%	Score 66.2	DB 6	Length 1890
Best Local Similarity	77.7%	Pred. No. 1.2e-27		
Matches 80: Conservative	0	Mismatches 23	Indels 0	Caps 0

```

RESULT 7
US-08-282-561-4
; Sequence 4, Application US/08282581
; Patent No. 5670349
; GENERAL INFORMATION:
; APPLICANT: Cramer, Carole L.
; APPLICANT: Weissenborn, Deborah L.
; TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM AND
; TITLE OF INVENTION: POST-HARVEST PRODUCTION OF GENE PRODUCTS
; TITLE OF INVENTION: IN PLANTS AND PLANT CELL CULTURES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

[illegible]

```

: TELEPHONE: 617/542-5070
:
: TELEFAX: 617/542-8906
:
: TELEX: 200154
:
: INFORMATION FOR SEQ ID NO:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 297 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: both
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA
:
: US-08-616-368A-17

```

Query Match 3.3%; Score 65.6; DB 1; Length 297;
Best Local Similarity 88.8%; Pred. No. 7.1e-08;
Matches 71; Conservative 0; Mismatches 9; Indels

Sequence	Conserved	Conservative	Mismatches	Indels	Gaps
Matches	71	71	0	9	0

[illegible]

RESULT 10
US-09-054-298-17
; Sequence 17, Application US/09054298

```

: Sequence 17, Application US/09054298
: Patent No. 6136553
: GENERAL INFORMATION:
: APPLICANT: Lee, Mu-En
: APPLICANT: Haber, Edgar
: APPLICANT: Jain, Mukesh
: APPLICANT: Yet, Shaw-Fang
: TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA

```

```
,  
, COUNTRY: USA  
,  
, ZIP: 02110-2804  
,  
, COMPUTER READABLE FORM:  
,  
, MEDIUM TYPE: Diskette  
,  
, COMPUTER: IBM Compatible  
,  
, OPERATING SYSTEM: Windows 95  
,  
, SOFTWARE: FastISO for Windows Version 2.0.b  
,  
, CURRENT APPLICATION DATA:  
,  
, APPLICATION NUMBER: US/09/054,298  
,  
, FILING DATE: 02-APR-1998  
,  
, PRIOR APPLICATION DATA:  
,  
, APPLICATION NUMBER: 08/616,368  
,
```

FILING DATE: 15-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Beattie, Ph.D., Ingrid A.
REGISTRATION NUMBER: P-42,306
REFERENCE/DOCKET NUMBER: 05433/022002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906

```

;          TELEX: 200154
;          INFORMATION FOR SEQ ID NO: 17:
;          SEQUENCE CHARACTERISTICS:
;             LENGTH: 297 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: both
;             TOPOLOGY: linear
;             MOLECULE TYPE: DNA

```

Query Match 3.3%; Score 65.6; DB 3; Length 297;
Best Local Similarity 88.8%; Pred. No. 7.1e-08;

[illegible]

```

RESULT 11
US-08-818-655-17
; Sequence 17, Application US/08818655
; Patent No. 6258557
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Haber, Edgar
; APPLICANT: Jain, Mukesh
; APPLICANT: Yet, Shuk-Fang
; TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 08/08/818,655
FILING DATE: 14-MAR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/616,368
FILING DATE: 15-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/030001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA

```

Query Match 3.3%; Score 65.6; DB 4; Length 297;
Best Local Similarity 88.8%; Pred. No. 7.1e-08;
Matches 71; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

[illegible]

RESULT 12
US-09-433-579-3/c
; Sequence 3, Application US/09433579

Search completed: December 16, 2002, 14:21:22
Job time : 161 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2002, 14:16:54 ; Search time 133 seconds
(without alignments)
5920.021 Million cell updates/sec

Title: US-09-848-852A-4
Perfect score: 1991
Sequence: 1 GCCTCCCTGTCGCCGACCC.....TGCTCTGGTCAATAAGATC 1991

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 35320 seqs, 197730502 residues 710640
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1991	100.0	1991	10	US-09-848-852A-4
2	442	22.2	503	10	Sequence 1710, Ap
3	368	18.5	432	10	Sequence 835, App
4	92	4.6	435	10	Sequence 4735, Ap
5	84.8	4.3	106	10	Sequence 4719, Ap
6	72.6	3.6	217	10	Sequence 1226, Ap
7	70.8	3.6	217	10	Sequence 1226, Ap
8	70.6	3.5	138	10	Sequence 792, App
9	70.6	3.5	169	10	Sequence 741, App
10	70.6	3.5	170	10	Sequence 780, App
11	70.6	3.5	242	10	Sequence 781, App
12	70.4	3.5	123	10	Sequence 182, App
13	70.4	3.5	202	10	Sequence 5, Appli
14	70.4	3.5	229	10	Sequence 1068, Ap
15	70	3.5	170	10	Sequence 82, Appl
16	70	3.5	209	10	Sequence 1107, Ap
17	69.4	3.5	144	10	Sequence 38, Appl
18	69.4	3.5	741	10	Sequence 479, App
19	69.2	3.5	212	10	Sequence 1097, Ap

C 20	69	3.5	419	10	US-09-969-373-237	Sequence 237, App
C 21	68.6	3.4	147	10	US-09-969-373-1517	Sequence 1517, Ap
C 22	68.6	3.4	252	10	US-09-969-373-1061	Sequence 1061, Ap
C 23	68.4	3.4	271	10	US-09-969-373-1313	Sequence 1313, Ap
C 24	68.4	3.4	313	10	US-09-969-373-1463	Sequence 1463, Ap
C 25	68	3.4	275	10	US-09-969-373-479	Sequence 479, App
C 26	68	3.4	276	10	US-09-969-373-478	Sequence 478, App
C 27	67.8	3.4	123	10	US-09-969-373-182	Sequence 182, App
C 28	67.8	3.4	225	10	US-09-969-373-622	Sequence 622, App
C 29	67.6	3.4	154	10	US-09-969-373-1052	Sequence 1052, Ap
C 30	67.6	3.4	170	10	US-09-969-373-82	Sequence 82, Appl
C 31	67.6	3.4	230	10	US-09-969-373-110	Sequence 110, App
C 32	67.6	3.4	230	10	US-09-969-373-111	Sequence 111, App
C 33	67.6	3.4	275	10	US-09-969-373-479	Sequence 479, App
C 34	67.6	3.4	276	10	US-09-969-373-478	Sequence 478, App
C 35	67.6	3.4	419	10	US-09-969-373-237	Sequence 237, App
C 36	67.4	3.4	147	10	US-09-969-373-1517	Sequence 1517, Ap
C 37	67.4	3.4	178	10	US-09-969-373-972	Sequence 972, App
C 38	67.4	3.4	205	10	US-09-969-373-1270	Sequence 1270, Ap
C 39	67.4	3.4	223	10	US-09-969-373-1183	Sequence 1183, Ap
C 40	67.2	3.4	238	10	US-09-969-373-1498	Sequence 1498, Ap
C 41	67.2	3.4	271	10	US-09-969-373-1313	Sequence 1313, Ap
C 42	67.2	3.4	453	10	US-09-867-701-5706	Sequence 5706, Ap
C 43	67	3.4	138	10	US-09-969-373-792	Sequence 792, App
C 44	67	3.4	229	10	US-09-969-373-1068	Sequence 1068, Ap
C 45	67	3.4	453	10	US-09-867-701-5706	Sequence 5706, Ap

ALIGNMENTS

RESULT 1
US-09-848-852A-4
; Sequence 4, Application US/09848852A
; Patent No. US20020106373A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; Tang, Y. Tom
; Corley, Neil C.
; Guegler, Karl J.
; Yue, Henry
; Patterson, Chandra
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH EGF SIGNALING
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/848,852A
; FILING DATE: 07-May-2001
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: 09/069,725
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0515 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1991 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: single									
; TOPOLOGY: linear									
; IMMEDIATE SOURCE:									
; LIBRARY: BEPINOT01									
; CLONE: 2056178									
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:									
US-09-848-852A-4									
Query Match 100.0%; Score 1991; DB 10; Length 1991;									
Best Local Similarity 100.0%; Pred No. 0;									
Matches 1991; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	GCCTCCCTGTC	CCCCGACCCTCTTTT	GATGCCCTCAGCAAGTGAAGAGGAAAGAGG	60				
DB	1	GCCTCCCTGTC	CCCCGACCCTCTTTT	GATGCCCTCAGCAAGTGAAGAGGAAAGAGG	60				
QY	61	AGGAGGAGGAGT	GAAGATGAAGAGGAGGAAGTGGCAGCTTGGAGCTGCCCCCAAGAT	120					
DB	61	AGGAGGAGGAGT	GAAGATGAAGAGGAGGAAGTGGCAGCTTGGAGCTGCCCCCAAGAT	120					
QY	121	GGAGTACGCTG	GGAACCTCCACGCGCCCGCCCTTCCCGCCCACTCATCGAAAAACCT	180					
DB	121	GGAGTACGCTG	GGAACCTCCACGCGCCCGCCCTTCCCGCCCACTCATCGAAAAACCT	180					
QY	181	GCTCACAGCC	CCCGCCGAGCCATGAGAGCCTTCCGGATGCTCTACTCAAAAAGCA	240					
DB	181	GCTCACAGCC	CCCGCCGAGCCATGAGAGCCTTCCGGATGCTCTACTCAAAAAGCA	240					
QY	241	CCTCGCTGAC	ATTCCACTGGAAGCTTTGGGGGGCGCCACCGGGCGGGCGGGGCTCG	300					
DB	241	CCTCGCTGAC	ATTCCACTGGAAGCTTTGGGGGGCGCCACCGGGCGGGCGGGGCTCG	300					
QY	301	CACACCCCA	AGAACCATCTTTACCCACGCAAGGGGGTGGACGCCACACAGTGCCATCCC	360					
DB	301	CACACCCCA	AGAACCATCTTTACCCACGCAAGGGGGTGGACGCCACACAGTGCCATCCC	360					
QY	361	CTGTGTGTG	CTTTTGACTCCCCCGGGGGCCACCTCCACCCGGCTGGTCTAGGTG	420					
DB	361	CTGTGTGTG	CTTTTGACTCCCCCGGGGGCCACCTCCACCCGGCTGGTCTAGGTG	420					
QY	421	CTCTATGCT	GAGGATGGGGTGAAGGCTCTCCACCACTGCGCCCTCTGGGCCCCCATGG	480					
DB	421	CTCTATGCT	GAGGATGGGGTGAAGGCTCTCCACCACTGCGCCCTCTGGGCCCCCATGG	480					
QY	481	AGGAAGATG	GACACTCAGGTGACCTCAAGTCTCCTCTGACCCCTGACTCGGGGCTCCTTT	540					
DB	481	AGGAAGATG	GACACTCAGGTGACCTCAAGTCTCCTCTGACCCCTGACTCGGGGCTCCTTT	540					
QY	541	CATGTACTC	TGCCCAACGGTTTGGGGGACAAATCTGGGCCAGAGGGAGCGCACTTGGC	600					
DB	541	CATGTACTC	TGCCCAACGGTTTGGGGGACAAATCTGGGCCAGAGGGAGCGCACTTGGC	600					
QY	601	ACCCCTGAT	GCCAGCATCCTCATCAGCATGTGTGCAGCATCGGGGACCATGTGGCCCA	660					
DB	601	ACCCCTGAT	GCCAGCATCCTCATCAGCATGTGTGCAGCATCGGGGACCATGTGGCCCA	660					
QY	661	GGAGCTTTT	TCAGGGCTCAGATTGGGCATGGCAGAAAGAGGACAGAGGCTCGGGAGAA	720					
DB	661	GGAGCTTTT	TCAGGGCTCAGATTGGGCATGGCAGAAAGAGGACAGAGGCTCGGGAGAA	720					
QY	721	AGCGGCCA	GACACAGCCCCCTGCGAGAGGAGCATGTGACCTGCGGTACAGAGCATCTTGA	780					
DB	721	AGCGGCCA	GACACAGCCCCCTGCGAGAGGAGCATGTGACCTGCGGTACAGAGCATCTTGA	780					
QY	781	CGAATTCCT	TCAACGATGGGAGCCTCATACCCCTCAGCACTGATGAGGTAGAGAA	840					
DB	781	CGAATTCCT	TCAACGATGGGAGCCTCATACCCCTCAGCACTGATGAGGTAGAGAA	840					
QY	841	GCTGGAGG	ACATTTTCCAGCAGAGTTTTCACCCCTCCAGGAAGGCTGGTGTGCA	900					
DB	841	GCTGGAGG	ACATTTTCCAGCAGAGTTTTCACCCCTCCAGGAAGGCTGGTGTGCA	900					
QY	901	GCTGATCC	AGTCTTACCAGCGGATGCCAGGCAATGCCATGGTGAGGGGCTTCGAGTGGC	960					

DB	901	GCTGATCC	AGTCTTACCAGCGGATGCCAGCAATGCGTGGGGCTTCGAGTGGC	960	
QY	961	TTATAAGC	GCACGTCTGACCATGATGACTTGGGAGCTTGTATGGACAGAACTGGCT	1020	
DB	961	TTATAAGC	GCACGTCTGACCATGATGACTTGGGAGCTTGTATGGACAGAACTGGCT	1020	
QY	1021	CAATGACC	AGGTGATGAACATGTATGAGACCTTGTATGGACACAGTCCCTGAAAAGGT	1080	
DB	1021	CAATGACC	AGGTGATGAACATGTATGAGACCTTGTATGGACACAGTCCCTGAAAAGGT	1080	
QY	1081	GCATTTCT	CAATAGTTCCTTATGATAAATCCGTACCAAGGTTATGATGGGTGAA	1140	
DB	1081	GCATTTCT	CAATAGTTCCTTATGATAAATCCGTACCAAGGTTATGATGGGTGAA	1140	
QY	1141	AAGGTGGC	CAAAAACGTGGACATCTTCAATAGGAGCTACTGCTAATCCCCATCCACCT	1200	
DB	1141	AAGGTGGC	CAAAAACGTGGACATCTTCAATAGGAGCTACTGCTAATCCCCATCCACCT	1200	
QY	1201	GAGGTGCA	TTGGTCCCTCATCTCTTGTATGTAGGCGACGCCACCATCACCTATTTTGA	1260	
DB	1201	GAGGTGCA	TTGGTCCCTCATCTCTTGTATGTAGGCGACGCCACCATCACCTATTTTGA	1260	
QY	1261	CTGCGACG	CTFACCTAAACCGCGCTGCCTAAGCATATTGCCAAGTATCTACAGGCAGA	1320	
DB	1261	CTGCGACG	CTFACCTAAACCGCGCTGCCTAAGCATATTGCCAAGTATCTACAGGCAGA	1320	
QY	1321	GCGGGTAA	AGAAAGACCGACTGGATTTCCACAGGCTGGAAGGTTACTTCAAAATGAA	1380	
DB	1321	GCGGGTAA	AGAAAGACCGACTGGATTTCCACAGGCTGGAAGGTTACTTCAAAATGAA	1380	
QY	1381	TGTGGCAG	GCGAGATAATGACAGTCACTGTGGTCTTTGTGCTTGCAGTACTGCAAGCA	1440	
DB	1381	TGTGGCAG	GCGAGATAATGACAGTCACTGTGGTCTTTGTGCTTGCAGTACTGCAAGCA	1440	
QY	1441	TCGTGGCC	CTGTCCAGCTTACGCCATTACCCAGCAGGACATGCCCAAATCTTGGTCGGCA	1500	
DB	1441	TCGTGGCC	CTGTCCAGCTTACGCCATTACCCAGCAGGACATGCCCAAATCTTGGTCGGCA	1500	
QY	1501	GATCTACA	AGGAGCTGTGCTCACTGCAAACTCACTGTGAGCCTCGTACCCAGACCCCA	1560	
DB	1501	GATCTACA	AGGAGCTGTGCTCACTGCAAACTCACTGTGAGCCTCGTACCCAGACCCCA	1560	
QY	1561	AGCCCAAA	TAATGGAGGAGACATGGGAGTCCCTTCCCAAGAAATCCAGTTCCCTTTC	1620	
DB	1561	AGCCCAAA	TAATGGAGGAGACATGGGAGTCCCTTCCCAAGAAATCCAGTTCCCTTTC	1620	
QY	1621	TCCTTGGC	CTTCCCACTCACTTCCCTTGGTTTTTTCATATTTAAATGTTTCAATTTCT	1680	
DB	1621	TCCTTGGC	CTTCCCACTCACTTCCCTTGGTTTTTTCATATTTAAATGTTTCAATTTCT	1680	
QY	1681	GTAATTTT	TTTTTCTTTGAGAGAACTTCTGATTTCTGATGTCAGGGGTGGCTACAG	1740	
DB	1681	GTAATTTT	TTTTTCTTTGAGAGAACTTCTGATTTCTGATGTCAGGGGTGGCTACAG	1740	
QY	1741	AAAAGCCC	CTTCTTCTGTTGAGGGAGTGTGGCCCTGTGCGCTGGGTGGAGCAG	1800	
DB	1741	AAAAGCCC	CTTCTTCTGTTGAGGGAGTGTGGCCCTGTGCGCTGGGTGGAGCAG	1800	
QY	1801	TCATCCTC	CCCCCTTCCCGTGCAGGAGCAGGAAATCACTGCTGGGGGTGGTGGCGGAC	1860	
DB	1801	TCATCCTC	CCCCCTTCCCGTGCAGGAGCAGGAAATCACTGCTGGGGGTGGTGGCGGAC	1860	
QY	1861	AATAGGAT	CACCTGCCCGAGATCTTCAACTTTTATATATATATATATATATATATA	1920	
DB	1861	AATAGGAT	CACCTGCCCGAGATCTTCAACTTTTATATATATATATATATATATATA	1920	
QY	1921	TATATATA	TATATATATATATATAAATATATAAATGCCAGGCTCCTGCTGCT	1980	
DB	1921	TATATATA	TATATATATATATATAAATATATAAATGCCAGGCTCCTGCTGCT	1980	
QY	1981	CAATAA	AGATC 1991		

; PRIOR FILING DATE: 2001-05-15

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2002, 15:26:39 ; Search time 2556 Seconds

(without alignments)
12615.498 Million cell updates/sec

Title: US-09-848-852A-4

Perfect score: 1991

Sequence: 1 GCCTCCCTGTCCTCCGACCC.....TGCTCTGGTCAATAAGATC 1991

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_Other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	874	43.9	1071	14	BM806659
2	812.2	40.8	1134	14	BM804530
3	805.6	40.5	1060	14	BM806502
c 4	786.4	39.5	817	9	AL581080
5	748.8	37.6	1114	13	BM476787
6	735	36.9	867	14	BQ440555

7	720.4	36.2	1099	13	BM555097
8	716.6	36.0	905	14	BM898865
c 9	679.2	34.1	994	14	BQ057204
10	676.8	34.0	936	14	BQ433698
c 11	673.8	33.8	715	12	BG255499
12	634.2	31.9	662	9	AL579669
13	629.4	31.6	676	10	AW961869
14	628.4	31.6	804	12	BG870520
15	626	31.4	626	14	BM696661
16	617.8	31.0	985	13	BI456217
17	594.2	29.8	790	13	BI079603
c 18	591.4	29.7	626	9	AL578220
c 19	590.6	29.7	657	14	BM976678
c 20	589.2	29.6	646	14	BQ021355
c 21	576.6	29.0	728	12	BG865703
22	576	28.9	886	14	BQ942601
23	562.4	28.2	665	14	BQ186996
24	560.6	28.2	1076	12	BF164921
c 25	554	27.8	607	14	BQ001101
c 26	542.8	27.3	605	10	AW085552
c 27	539	27.1	835	13	BI732013
28	536	26.9	704	10	AV721960
29	532.6	26.8	620	10	AW578424
30	532.6	26.8	620	10	AW604359
31	523	26.3	552	10	AW964755
c 32	522.6	26.2	667	17	AZ399027
c 33	515.6	25.9	567	9	AI800802
34	509	25.6	689	12	BF608506
c 35	503.8	25.3	616	10	AW916790
c 36	503	25.3	1070	13	BI247525
37	498.2	25.0	643	14	N36589
c 38	493.4	24.8	529	12	BF195133
c 39	491	24.7	508	13	BM670002
c 40	485.6	24.4	599	13	BI017569
41	485.2	24.4	555	14	BQ220291
42	481.8	24.2	549	13	BM507574
43	478.6	24.0	819	12	BF139554
44	475.4	23.9	504	12	BF954315
45	474	23.8	632	12	BG083248

ALIGNMENTS

RESULT 1	BM806659	BM806659	1071 bp	mRNA	linear	EST 05-MAR-2002	
LOCUS	AGENCOURT_6542888	NIH_MGC_72	Homo sapiens	cDNA clone	IMAGE:5548958		
DEFINITION	5', mRNA sequence.						
ACCESSION	BM806659						
VERSION	BM806659.1	GI:19123482					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1071) NIH-MGC http://mgc.mci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC/DCTD/DTF CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM12258 row: 1 column: 15 High quality sequence stop: 648. Location/Qualifiers 1. .1071					
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
COMMENT							
FEATURES							
SOURCE							

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5548958"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT      255 a      264 c      271 g      278 t      3 others
ORIGIN
Query Match      43.9%; Score 874; DB 14; Length 1071;
Best Local Similarity 96.5%; Pred. No. 3.5e-154;
Matches 947; Conservative 0; Mismatches 26; Indels 8; Gaps 5;
QY 742 GCGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGGAGCAATCTCTCAAACGATGCG 801
DB 1 GCGAGAGGAGCATGTGACCTGCGTACAGAGCATCTTGGAGCAATCTCTCAAACGATGCG 60
QY 802 CAGCCTCATACCCCTCAGACATGATGAGTAGAGAGCTGAGAGCATTTTCCAGCA 861
DB 61 CAGCCTCATACCCCTCAGACATGATGAGTAGAGAGCTGAGAGCATTTTCCAGCA 120
QY 862 GGAGTTTTCACCCCTTCCAGGAAGGCGCTGGTGTTCAGCTGATCCAGTCTTACCAGCG 921
DB 121 GGAGTTTTCACCCCTTCCAGGAAGGCGCTGGTGTTCAGCTGATCCAGTCTTACCAGCG 180
QY 922 GATGCCAGGCAATGCCATGTTGAGGGCTTCCGAGTGGCTTATAAGCGGCACGTCTGAC 981
DB 181 GATGCCAGGCAATGCCATGTTGAGGGCTTCCGAGTGGCTTATAAGCGGCACGTCTGAC 240
QY 982 CATGGATGACTTGGGACCTTGTATGACAGAACTGCTCAATGACAGGTGATGAACAT 1041
DB 241 CATGGATGACTTGGGACCTTGTATGACAGAACTGCTCAATGACAGGTGATGAACAT 300
QY 1042 GTATGAGACCTGGTATGACACAGTCCCTGAAAGGTGCAATTTCTCAATAGTTTCT 1101
DB 301 GTATGAGACCTGGTATGACACAGTCCCTGAAAGGTGCAATTTCTCAATAGTTTCT 360
QY 1102 CTATGATAACTCCGTACCAAGGTTATGATGGGTGAAAAGTGGACCAAAACGTGGA 1161
DB 361 CTATGATAACTCCGTACCAAGGTTATGATGGGTGAAAAGTGGACCAAAACGTGGA 420
QY 1162 CATCTTCAATAAGAGCTACTGCTAATCCCATCCACCTGGAGGTGCAATGGTCCCTCAT 1221
DB 421 CATCTTCAATAAGAGCTACTGCTAATCCCATCCACCTGGAGGTGCAATGGTCCCTCAT 480
QY 1222 CTCTGTTGATGTGAGCGAGCGACCATCATCTATTTTGACTCGCAGCGTACCCCTAAACCG 1281
DB 481 CTCTGTTGATGTGAGCGAGCGACCATCATCTATTTTGACTCGCAGCGTACCCCTAAACCG 540
QY 1282 CCCTGCGCTTAAGCATATGCCCAGTATCTACAGGAGAGGCGGTAAAGAACACCGACT 1341
DB 541 CCCTGCGCTTAAGCATATGCCCAGTATCTACAGGAGAGGCGGTAAAGAACACCGACT 600
QY 1342 GGATTTCCACAGGCGTGGAAAGGTTACTTCAAATGAATGTGCCAGCAGCAATAATGA 1401
DB 601 GGATTTCCACAGGCGTGGAAAGGTTACTTCAAATGAATGTGCCAGCAGCAATAATGA 660
QY 1402 CAGTGACTGTGGTGTGTTGTGTGAGTACTGCAAGCATCTGCGCCCTGTCTCAGCCATT 1461
DB 661 CAGTGACTGTGGTGTGTTGTGTGAGTACTGCAAGCATCTGCGCCCTGTCTCAGCCATT 720
QY 1462 CAGCTTCAACCCAGCAGCATGCCAAACTTCGTGGGAGATCTACAAGAGTGTGTCA 1521
DB 721 CAGCTTCAACCCAGCAGCATGCCAAACTTCGTGGGAGATCTACAAGAGTGTGTCA 780
QY 1522 CT-GCAAACTCACTGTGTGAGCTCTGTACCCAGAGCCCAAGCCCAATAATGGGAAGGA 1580
DB 781 CTGGCAAACTCACTGTGTGAGCTCTGTACCCAGAGCCCAAGCCCAATAATGGGAAGGA 839
```

```
QY 1581 GACATGGGAGTCCCTTCCCAAGAAACTCCAGTT-CCTTTCCTCTCTTGCTCTTCCACT 1639
DB 840 GACATGGGAGTCCCTTCCCAAGAAACTCAGTTCCTCTCTCTCTCTCTCTCTCTCTCT 899
QY 1640 CACTTCCC---TTTGTTTTCATATTTAAAT--GTTTCAATTTCTGTAATTTTTTCT 1694
DB 900 CACTTTCCTCTTTTGGTTTTCATATTTAAATGGTTCATATTTCCCTGGATTTTTTTC 959
QY 1695 TTGAGAGAAATACATGTTGATT 1715
DB 960 CTTTGAGAAAGATACITGGNT 980

RESULT 2
BM804530      1134 bp      mRNA      linear      EST 05-MAR-2002
LOCUS      AGENCOURT_6492923 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5554637
DEFINITION      5', mRNA sequence.
ACCESSION      BM804530
VERSION      BM804530.1 GI:19121353
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1134)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-r@mail.nih.gov
      Tissue Procurement: Lou Staudt
      cDNA Library Preparation: Life Technologies, Inc.
      DNA Sequencing by: Agencourt Bioscience Corporation
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLAM12273 row: 1 column: 06
      High quality sequence stop: 583.

FEATURES
      source
      1..1134
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:5554637"
      /clone_lib="NIH_MGC_85"
      /tissue_type="lymphoma, cell line"
      /lab_host="DH10B (phage-resistant)"
      /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
      Site_2: SalI; Cloned unidirectionally; Oligo-dt primed.
      Average insert size 1.867 kb. Library enriched for
      full-length clones and constructed by Life Technologies.
      Note: this is a NIH_MGC Library."

BASE COUNT      261 a      326 c      311 g      236 t
ORIGIN
Query Match      40.8%; Score 812.2; DB 14; Length 1134;
Best Local Similarity 96.1%; Pred. No. 1.4e-142;
Matches 854; Conservative 0; Mismatches 33; Indels 2; Gaps 2;
QY 286 GCGCGCGGCGCTCGCACACCCCAAGAACATCTTTCACCCCAAGAGGGGTGCGACGC 345
DB 1 GGACGCGTGGTCCGACACACCCCAAGAACATCTTTCACCCCAAGAGGGGTGCGACGC 60
QY 346 CACAGGTGCCATCCCTCTGTCGTTTTCCTCCCGGGGGCCACCTCCACCCCGCG 405
DB 61 CACAGGTGCCATCCCTCTGTCGTTTTCCTCCCGGGGGCCACCTCCACCCCGCG 120
QY 406 TGGGTCGTAGTGTCTCTCATGGTGGAGTGGGTGAGAGGTCTCCACCAAGTCCCT 465
DB 121 TGGGTCGTAGTGTCTCTCATGGTGGAGTGGGTGAGAGGTCTCCACCAAGTCCCT 180
QY 466 CTGGGCCCCCCCCATGGAGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTGGACCCCTG 525
```

```
Db 181 CTGGGCCCCATGAGGAAGATGACTCAGGTGACTCCAAAGTCTCTCTGGACCCCTG 240
QY 526 ACTGGGGCTCTTTCATGTACTCTGCCCAACGGTTTGGGGGACAATCTGGGCCAGAA 585
Db 241 ACTGGGGCTCTTTCATGTACTCTGCCCAACGGTTTGGGGGACAATCTGGGCCAGAA 300
QY 586 GGGAGCGCA-CATTGGCACCCCTGATGCGCAGCATCTCATAGCAATGTGTGCAGCATCG 644
Db 301 GGGAGCGCAGCTTGCACCCCTGATGCGCAGCATCTCATAGCAATGTGTGCAGCATCG 360
QY 645 GGGACCATGTGGCCGAGGAGCTTTTCAGGGCTCAGATTGGGCTAGAGTGGCAGAGGAGG 704
Db 361 GGGACCATGTGGCCGAGGAGCTTTTCAGGGCTCAGATTGGGCTAGAGTGGCAGAGGAGG 420
QY 705 AGAGGCTGGGAGAAAGCGGCCAGCACAGCCCTCGGAGAGGAGCATGTGACCTGG 764
Db 421 AGAGGCTGGGAGAAAGCGGCCAGCACAGCCCTCGGAGAGGAGCATGTGACCTGG 480
QY 765 TACAGAGCATCTTGGACGAATTCCTTCAAAAGTATGGCAGCCTCATACCCCTCAGCACTG 824
Db 481 TACAGAGCATCTTGGACGAATTCCTTCAAAAGTATGGCAGCCTCATACCCCTCAGCACTG 540
QY 825 ATGAGGTAGTAGAAGCTGGAGGACATTTTCCAGCAGAGTFTTCCACCCCTTCCAGGA 884
Db 541 ATGAGGTAGTAGAAGCTGGAGGACATTTTCCAGCAGAGTFTTCCACCCCTTCCAGGA 600
QY 885 AGGGCTGGTGTGCGAGCTGATCCAGTCTTACCAGCGATGCCAGGCAATGCCATGGTGA 944
Db 601 AGGGCTGGTGTGCGAGCTGATCCAGTCTTACCAGCGATGCCAGGCAATGCCATGGTGA 660
QY 945 GGGGCTCCGAGTGGCTTATAAGCGGCACGCTGTGACATGGATGACTTGGGGACCTTGT 1004
Db 661 GGGGCTCCGAGTGGCTTATAAGCGGCACGCTGTGACATGGATGACTTGGGGACCTTGT 720
QY 1005 ATGGACAGAACTGGCTCAATGACAGGTGATGAACATGTATGGAGACCTTGTATGGACA 1064
Db 721 ATGGACAGAACTGGCTCAATGACAGGTGATGAACATGTATGGAGACCTTGTATGGACA 780
QY 1065 CAGTCCCTGAAAAGTGTATTTCTCAATAGTTTCTTATGATATAAATCCGTACCAAG 1124
Db 781 CA-TCCCTGAAAAGTGTATTTCTCAATAGTTTCTTATGATATAAATCCGTACCAAG 839
QY 1125 GTTATGATGGGTGAAAAGTGGACAAAACGTGGACATCTTCAATAA 1173
Db 840 GTTATGAGGGGTGAAAAGTGGACAAAAGGGGAAACCTCTTTCA 888
```

```
RESULT 3
LOCUS BM806502
DEFINITION AGENCOURT_6542932 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5548865
5', mRNA sequence.
ACCESSION BM806502
VERSION BM806502.1 GI:19123325
KEYWORDS EST.
SOURCE human.
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
1 (bases 1 to 1060)
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

```
Plate: LLAM12258 row: h column: 18
High quality sequence stop: 652.
FEATURES
source
Location/Qualifiers
1..1060
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5548865"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 248 a 272 c 304 g 236 t
ORIGIN
```

```
Query Match 40.5%; Score 805.6; DB 14; Length 1060;
Best Local Similarity 95.8%; Pred. No. 2.4e-141;
Matches 883; Conservative 0; Mismatches 29; Indels 10; Gaps 5;
QY 503 TCCAAAGTCTCTCTGGACCCCTGACTCGGGCTCCTTTTCATCTACTCTGCCCAACGGTTT 562
Db 1 TCCAAAGTCTCTCTGGACCCCTGACTCGGGCTCCTTTTCATCTACTCTGCCCAACGGTTT 60
QY 563 TGGGGACAATCTGGGCCAGAAAGGGAGGCGCA-CCTGGCACCCCTGTATGCCAGCATCCT 621
Db 61 TGGGGACAATCTGGGCCAGAAAGGGAGGCGCAGCTTGGCACCCCTGTATGCCAGCATCCT 120
QY 622 CATCAGCAATGTGTGAGCATCGGGACCATGTGCCCCAGGAGCTTTTTCAGGGCTCAGA 681
Db 121 CATCAGCAATGTGTGAGCATCGGGACCATGTGCCCCAGGAGCTTTTTCAGGGCTCAGA 180
QY 682 TTTGGGCATGGCAGAAAGGAGAGAGGCTTGGGAGAAAGCGGCCACACAGCCCCCT 741
Db 181 TTTGGGCATGGCAGAAAGGAGAGAGGCTTGGGAGAAAGCGGCCACACAGCCCCCT 240
QY 742 GCGAGAGGAGCATGTGACCTGCTACAGAGCATCTTGGAGCAATCTCTTCAACAGTATGG 801
Db 241 GCGAGAGGAGCATGTGACCTGCTACAGAGCATCTTGGAGCAATCTCTTCAACAGTATGG 300
QY 802 CAGCCTCATACCCCTCAGCACTGATGAGTAGTAGAAGAGCTGGAGGACATTTTCCAGCA 861
Db 301 CAGCCTCATACCCCTCAGCACTGATGAGTAGTAGAAGAGCTGGAGGACATTTTCCAGCA 360
QY 862 GGAGTTTTCACCCCTTCCAGGAAGGGCTGTGTGAGCTGATCCAGCTCTTACCAGCG 921
Db 361 GGAGTTTTCACCCCTTCCAGGAAGGGCTGTGTGAGCTGATCCAGCTCTTACCAGCG 420
QY 922 GATGCCAGGCAATGCATGGTCAGGGGCTTCCGAGTGGCTTATAAGCGGCAGCTGTGAC 981
Db 421 GATGCCAGGCAATGCATGGTCAGGGGCTTCCGAGTGGCTTATAAGCGGCAGCTGTGAC 480
QY 982 CATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCTCAATGACAGGATGATGAACAT 1041
Db 481 CATGGATGACTTGGGGACCTTGTATGGACAGAACTGGCTCAATGACAGGATGATGAACAT 540
QY 1042 GTATGAGACCTTGGTCATGCGACACAGTCCCTGAAAAGGTGCAATTTCTCAATAGTTTCT 1101
Db 541 GTATGAGACCTTGGTCATGCGACACAGTCCCTGAAAAGGTGCAATTTCTCAATAGTTTCT 600
QY 1102 CTATGATAAATCCGTACCAGGGTTATGATGGGTGAAAAGGTGACCAAAACCTGGA 1161
Db 601 CTATGATAAATCCGTACCAGGGTTATGATGGGTGAAAAGGTGACCAAAACCTGGA 660
QY 1162 CATCTTTCAATAAGGAGCTACTGCTAATCCCATCCACCTGGAGGTCATTGGTCCCTCAT 1221
Db 661 CATCTTTCAATAAGGAGCTACTGCTAAT -CCCATCCACCTGGAGGTCATTGGTCCCTCAT 719
QY 1222 CTCCTGTGTATGAGGCGAGCAGCACCATCACCTATTTTGTACTGCGACGGTACCCCTAAACG 1281
Db 720 CTCCTGTGTATGAGGCGAGCAGCACCATCACCTATTTTGTACTGCGACGGTACCCCTAGACG 779
```

QY 1282 CCGCTGCCCTAAGCATATGCCAAGTATCTACAGGCAGA--GGCGGTAAGAAAGACCGCA 1339
 Db 780 CCGCTGCCCTAAGCATATGCCAAGTATCTACAGGCAGAAGCGGTTAAGAAAGACCGCA 839
 QY 1340 CTGGATTT---CCACCAGGCGTGAAGAGTTACTTCAAAATGAATG--TGGCCAGGCAG 1393
 Db 840 CTGGATTTCCACACACCGGCTGGAAAGGTTACTTCAAAATGAATGTTGGCCAGGCAAG 899
 QY 1394 AATAATCACAGTGAAGTCTGGTG 1415
 Db 900 AATAATCACAGTGACCGGTGG 921

RESULT 4
 AL581080/c 817 bp mRNA linear EST 16-FEB-2001
 LOCUS AL581080 LRI_NFL008_Tc2 Homo sapiens cDNA clone CS0DJ013YK06 3
 DEFINITION prime, mRNA sequence.
 ACCESSION AL581080
 VERSION AL581080
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 817)
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1..817
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DJ013YK06"
 /clone_lib="LTI_NFL008_Tc2"
 /sex="male"
 /tissue_type="T cells from T cell leukemia"
 /notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 217 a 199 c 206 g 190 t 5 others
 ORIGIN

Query Match 39.5%; Score 786.4; DB 9; Length 817;
 Best Local Similarity 98.9%; Pred. No. 1e-137;
 Matches 809; Conservative 4; Mismatches 1; Indels 4; Gaps 2;

QY 1075 AAGGTGCATTTCTTCAATAGTTCTCTATGATATAACCCGTAACAGGGTTATGATGG 1134
 Db 817 AAGGTGCATTTCTTCAATA-KTTCTCTATGATAAATCCGTAACAGGGTTATGATGG 759
 QY 1135 GGTGAAGGTGACCAAAACGTGGACATCTTCAATAAGGAGCTACTGTAATCCCAT 1194
 Db 758 GGTGAAGGTGACCAAAACGTGGACATCTTCAATAAGGAGCTACTGTAATCCCAT 699
 QY 1195 CCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTGAGCGGACGACCATCACCTA 1254
 Db 698 CCACCTGGAGGTGCATTTGGTCCCTCATCTCTGTTGATGTGAGCGGACGACCATCACCTA 639
 QY 1255 TTTTGACTCCAGCGTACCCCTAAACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACA 1314
 Db 638 TTTTGACTCCAGCGTACCCCTAAACCGCGCTGCCCTAAGCATATTGCCAAGTATCTACA 579

QY 1315 GGCAGAGCGGTAAAGAAAGACCGACTGGATTTCCACCAGGCTGGAAGGTTACTTCAA 1374
 Db 578 GGCAGAGCGGTAAAGAAAGACCGACTGGATTTCCACCAGGCTGGAAGGTTACTTCAA 519
 QY 1375 AATGAATGTGGCGAGCGAGCAATAATGACAGTACTGTGGTGTCTTTGTGTGTCAGTACTG 1434
 Db 518 AATGAATGTGGCGAGCGAGCAATAATGACAGTACTGTGGTGTCTTTGTGTGTCAGTACTG 459
 QY 1435 CAAGCATCTGGCCCTGTCTCAGCCATTTCAGCTTCAACCAGCAGGACATGCCAAACTTCG 1494
 Db 458 CAAGCATCTGGCCCTGTCTCAGCCATTTCAGCTTCAACCAGCAGGACATGCCAAACTTCG 399
 QY 1495 TCGGCAGATCTCAAGAGAGCTGTCTCACTCAAACTCACTGTCTGAGCCCTCGTACCCAG 1554
 Db 398 TCGGCAGATCTCAAGAGAGCTGTCTCACTCAAACTCACTGTCTGAGCCCTCGTACCCAG 339
 QY 1555 ACCCAAGCCCCATAAATGGAGGAGGACATGGAGTCCCTTCCCAAGAAATCCAGTTC 1614
 Db 338 ACCCAAGCCCCATAAATGGAGGAGGACATGGAGTCCCTTCCCAAGAAATCCAGTTC 279
 QY 1615 CTTTCTCTCTGCCCTTCCCACTCACTTCCCTTTGGTGTTCATATTTAAATGTTCA 1674
 Db 278 CTTTCTCTCTGCCCTTCCCACTCACTTCCCTTTGGTGTTCATATTTAAATGTTCA 219
 QY 1675 ATTCTGTATTTTCTTTCTTTGAGAGAACTACTTGTGATTTCTGATGTGCAGGGGGTGG 1734
 Db 218 ATTCTGTATTTTCTTTCTTTGAGAGAACTACTTGTGATTTCTGATGTGCAGGGGGTGG 159
 QY 1735 CTACAGAAAAGCCCCCTTCTCTCTGTTTGGAGGAGTGTGGCCCTGTGGCTGGGTG 1794
 Db 158 CTACAGAAAAGCCCCCTTCTCTCTGTTTGGAGGAGTGTGGCCCTGTGGCTGGGTG 99
 QY 1795 GACGAGTCACTCTCC---CCTTCCCGTGCAGGAGCAGGAATCAGTCTGGGGTGG 1851
 Db 98 GACGAGTCACTCTSSCGCKGTTCCCGTGCAGGAGCAGGAATCAGTCTGGGGTGG 39
 QY 1852 TGGCGCGACAATAAGGATCACTGCTGCCAGATCTTCAA 1889
 Db 38 TGGCGCGACAATAAGGATCACTGCTGCCAGATCTTCAH 1

RESULT 5
 BM476787
 LOCUS BM476787
 DEFINITION 5', mRNA sequence.
 ACCESSION BM476787
 VERSION BM476787.1 GI:18525829
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1114)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12275 row: f column: 03
 High quality sequence stop: 554.
 Location/Qualifiers
 1..1114
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

RESULT 6	BQ440555	867 bp	linear	EST 24-MAY-2002
LOCUS	AGENCOURT_7827074 NIH_MGC_67	867 bp	linear	EST 24-MAY-2002
DEFINITION	5', mRNA sequence.			
ACCESSION	BQ440555			
VERSION	BQ440555.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 867)			
TITLE	NIH-MGC http://mgc.nci.nih.gov/.			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: rcgabs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM13491 row: 0 column: 20 High quality sequence stop: 645. Location/Qualifiers 1. 867 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6153067" /clone_lib="NIH_MGC_67" /tissue_type="retinoblastoma" /lab_host="DH10B (phage-resistant)" /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies." a 266 c 261 g 148 t 1 others			
BASE COUNT	191 a 266 c 261 g 148 t 1 others			
ORIGIN				
Query Match	36.9%	Score 735;	DB 14;	Length 867;
Best Local Similarity	98.2%	Pred. No. 4.5e-128;		
Matches	785; Conservative	0; Mismatches 10; Indels	4; Gaps	4;
Qy	30	GCCTCAGCAAGTGAAGAGGAGGAAGAGAGGAGGAGGAGGAGTGAAGATGAAGAGGAG 89		
Db	12	GCCTCAGCAAGTGAAGAGGAGGAAGAGAGGAGGAGGAGGAGTGAAGATGAAGAGGAG 71		
Qy	90	GAAGTGGCAGTTGGAGGCTGCCCCCAAGATGGAGTCACTGCTGGGAACCTCCAGCGGGCC 149		
Db	72	GAAGTGGCAGTTGGAGGCTGCCCCCAAGATGGAGTCACTGCTGGGAACCTCCAGCGGGCC 131		
Qy	150	CGCCCTTCCGCCCACTCATCGAAAACCTGCTCACAGCGCCGCCCGAGCCATGAGA 209		
Db	132	CGCCCTTCCGCCCACTCATCGAAAACCTGCTCACAGCGCCGCCCGAGCCATGAGA 191		
Qy	210	GCCTTCGGATGCTGCTCTACTCAAAAAGCACCTCGCTGCATCTCCACTGGAAGCTTTGG 269		
Db	192	GCCTTCGGATGCTGCTCTACTCAAAAAGCACCTCGCTGCATCTCCACTGGAAGCTTTGG 251		
Qy	270	GGGGCCACCGGGCCCGGGCGGGGCGCTCGCACACCCCAAGAACCATCTTTACCCACG 329		
Db	252	GGGGCCACCGGGCCCGGGCGGGGCGCTCGCACACCCCAAGAACCATCTTTACCCACG 311		
Qy	330	CAAGGGGGTGGAGCCGCACAGGTGCCATCCCTCTGTTGTCGTTTTGACTCCCCCGGGG 389		
Db	312	CAAGGGGGTGGAGCCGCACAGGTGCCATCCCTCTGTTGTCGTTTTGACTCCCCCGGGG 371		
Qy	390	CCAGCTCACACCCGGGCTGGTCTGCTAGGTGCTCTCATGCTGAGGTGGGCTGAGAGG 449		

TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: L7AM13564 row: o column: 14
High quality sequence stop: 614.

FEATURES

Location/Qualifiers

source

1. 905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6181093"
/clone_lib="Lupski_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"

/note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCCG-3' and
5'-GACTAGTCTAGATCGCAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."

BASE COUNT 212 a 238 c 220 g 235 t
ORIGIN

Query Match 36.0%; Score 716.6; DB 14; Length 905;
Best Local Similarity 96.1%; Pred. No. 1.3e-124;
Matches 756; Conservative 0; Mismatches 29; Indels 2; Gaps 2;

QY 946 GGGCTTCGAGTGGGTTATAGCGCAGCGTCTGCTACCATGGATGGAGCTTGGGACCTGTGA 1005*
|||||
Db 1 GGGCTTCGAGTGGGTTATAGCGCAGCGTCTGCTACCATGGATGGAGCTTGGGACCTGTGA 60
|||||

QY 1006 TGSACAGACTGGCTCAATGACAGGTGATGACATGATGAGACCTGGTCTATGGACAC 1065
|||||

Db 61 TGGACAGAACTGGCTCAATGACAGGTGATGAACATGATGGAGACCTGGTCTATGGACAC 120
|||||

QY 1066 AGTCCTGAAAAAGGTGCATTTCTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGG 1125
|||||

Db 121 AGTCCTGAAAAAGGTGCATTTCTCAATAGTTTCTTCTATGATAAACTCCGTACCAAGGG 180
|||||

QY 1126 TTATGATGGGTGAAAAGGTGGACCAAAACCTGGACATCTTCAATAAGAGGACTACTGCT 1185
|||||

Db 181 TTATGATGGGTGAAAAGGTGGACCAAAACCTGGACATCTTCAATAAGAGGACTACTGCT 240
|||||

QY 1186 ATATCCCATCCACCTGGAGGTGCATGTGTCCTCATCTCTGTGTATGATGAGCGGACGAC 1245
|||||

Db 241 AATCCCATCCACCTGGAGGTGCATGTGTCCTCATCTCTGTGTATGATGAGCGGACGAC 300
|||||

QY 1246 CATCACTATTGTTGACTCGCAGCGTACCTTAACCGCGCTGCCCTTAAGCATATTGCCAA 1305
|||||

Db 301 CATCACTATTGTTGACTCGCAGCGTACCTTAACCGCGCTGCCCTTAAGCATATTGCCAA 360
|||||

QY 1306 GTATCTACAGGCAGAGCGGTAAAGAAAGACCGACTGGATTTCCACCGGGCTGAAAGG 1365
|||||

Db 361 GTATCTACAGGCAGAGCGGTAAAGAAAGACCGACTGGATTTCCACCGGGCTGAAAGG 420
|||||

QY 1366 TTACTTCAAAATGAATGGCCAGGCAGCAATTAATGACAGTGAAGTGGTCTTTTGTGTT 1425
|||||

Db 421 TTACTTCAAAATGAATGGCCAGGCAGCAATTAATGACAGTGAAGTGGTCTTTTGTGTT 480
|||||

QY 1426 GCAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTACAGCTTCACCCAGCAGCATGCC 1485
|||||

Db 481 GCAGTACTGCAAGCATCTGGCCCTGTCTCAGCCATTACAGCTTCACCCAGCAGCATGCC 1540
|||||

QY 1486 CAAACTTCGTCGGCAGATCTACAAGAGCTGTGTCTCACTCAAACTCACTGTGTGAGCCTC 1545
|||||

Db 541 CAAACTTCGTCGGCAGATCTACAAGAGCTGTGTCTCACTCAAACTCACTGTGTGAGCCTC 600
|||||

QY 1546 GTACCCCCAGA-CCCCAAGCCCAATAATGGGAAGGAGAGCATGGAGTCCCTTCCCAAGAA 1604
|||||

Db 601 GTACCCCCAGACCCCAAGCCCAATAATGGGAAGGAGAGCATGGAGTCCCTTCCCAAGAA 560
|||||

QY 1605 ACTCCAGTTCCTTCT 1663
|||||

Db 661 ACTCCAGTTCCTTCT 720
|||||

QY 1664 TAAATGTTTCAATTTCTGTATTTTTTTTTTCTTTTGAGAGAACTGTTGATTTCTGATCT 1723
|||||

Db 721 TAAATGTTTCAATTTCTGGATTTTTTTTTTCTTTGACGAGAATACCCGGTTTCATTTCT 780
|||||

QY 1724 GCAGGGG 1730
|||||

Db 781 GAATGGG 787
|||||

RESULT 9

LOCUS BQ057204/c

DEFINITION BQ057204 994 bp mRNA linear EST 29-MAR-2002
5', mRNA sequence.
AGENCOURT_6769328 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5812399

ACCESSION BQ057204

VERSION BQ057204.1 GI:19816544

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 994)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Lou Staudt

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L7COM2062 row: m column: 08

High quality sequence stop: 620.

Location/Qualifiers

source

1. 994

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5812399"

/clone_lib="NIH_MGC_99"

/tissue_type="lymphoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 278 a 228 c 232 g 256 t

ORIGIN

Query Match

Best Local Similarity 34.1%; Score 679.2; DB 14; Length 994;

Pred. No. 1.3e-117;

Matches	720;	Conservative	0;	Mismatches	18;	Indels	20;	Gaps	1;
QY	1231	TGTGAGCGGACGACCATACCTATTTTGTGCTGGCAGCGTACCTAAACCGCGCTCC	1290						
Db	795	TGAGGGGACGCGCCCATCTTATTTGACTCGCAGGGTCCCTTAACCGCGCTCC	736						
QY	1291	TAAGCATATTGCCAAGTATCTACAGGCGAGGGGTAAAGAAACCGACTGATTTCCA	1350						
Db	735	TAAGCATATTGCCAAGTATCTACAGGCGAGGGGTAAAGAAACCGACTGATTTCC	676						
QY	1351	CCAGGGCTGGAAGTTTACTTCAAAATGAATGTGGCCAGGAGCAATACAGTACTG	1410						
Db	675	CCAGGGCTGGAAGTTTACTTCAAAATGAATGTGGCCAGGAGCAATACAGTACTG	616						
QY	1411	TGTTGCTTTTGTGTGAGTACTGCAAGCATCTGGCCCTCTCTCAGCCATTCAGCTTCA	1470						
Db	615	TGTTGCTTTTGTGTGAGTACTGCAAGCATCTGGCCCTCTCTCAGCCATTCAGCTTCA	556						
QY	1471	CCAGCAGGACATGCCAAACTTCGTGGCGAGATCTACAAGGAGTGTGTCACTGCAAACT	1530						
Db	555	CCAGCAGGACATGCCAAACTTCGTGGCGAGATCTACAAGGAGTGTGTCACTGCAAACT	496						
QY	1531	CACGTGTGAGCTCTGTACCCAGACCCCAAGCCCATAAATGGAGGGAGACATGGGAG	1590						
Db	495	CACGTGTGAGCTCTGTACCCAGACCCCAAGCCCATAAATGGAGGGAGACATGGGAG	436						
QY	1591	TCCTTCCCAAGAAACGCCAGTTTCTTCTCTCTGCTCTTCCCTTCCCACTCACCTCCCTT	1650						
Db	435	TCCTTCCCAAGAAACGCCAGTTTCTTCTCTCTGCTCTTCCCTTCCCACTCACCTCCCTT	376						
QY	1651	GGTTTTTCATATTAATGTTTCAATTTCTGTATTTTCTTTTCTTGAGAGATACTTGT	1710						
Db	375	GGTTTTTCATATTAATGTTTCAATTTCTGTATTTTCTTTTCTTGAGAGATACTTGT	316						
QY	1711	TGATTTCTGATGTCAGGGGGTGGCTACAGAAAGCCCTTCTCTCTGTTTGGAGGG	1770						
Db	315	TGATTTTGTATGTCAGGGGGTGGCTACAGAAAGCCCTTCTCTCTGTTTGGAGGG	256						
QY	1771	GAGTGTGGCCCTGTGGCTGGGTGGAGCAGTATCTCTCCCTTCCCGTGCAGGAGCA	1830						
Db	255	GAGTGTGGCCCTGTGGCTGGGTGGAGCAGTATCTCTCCCTTCCCGTGCAGGAGCA	196						
QY	1831	GGAATCAGTGTGGGGGTGGGGGGGACAAATAGATCACTCCCTGCCAGATCTTCAA	1890						
Db	195	GGAATCAGTGTGGGGGTGGGGGGGACAAATAGATCACTCCCTGCCAGATCTTCAA	136						
QY	1891	CTTTA	1950						
Db	135	CTTT-----TA	96						
QY	1951	AATATATAATGCACGGTCTCTCTGGTCAATAAG	1988						
Db	95	AATATATAATGCACGGTCTCTCTGGTCAATAAG	58						

RESULT 10
LOCUS BQ433698
DEFINITION AGENCOURT_7770396 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6025493
5', mRNA sequence.
ACCESSION BQ433698
VERSION BQ433698.1 GI:21172774
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 936)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13237 row: 1 column: 06
High quality sequence start: 41
High quality sequence stop: 300.

FEATURES
source
I. .936
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6025493"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life Technologies."

BASE COUNT 200 a 296 c 283 g 157 t
ORIGIN
Query Match 34.0%; Score 676.8; DB 14; Length 936;
Best Local Similarity 96.0%; Pred. No. 3.8e-117;
Matches 726; Conservative 0; Mismatches 27; Indels 3; Gaps 3;

QY 1 GCCTCCTGTCCCGGACCCCTCTTTTGTGCTCAGCAAGTGAAGAGGAGGAGAGAGG 60
Db 87 GCCTCCTGTCCCGGACCCCTCTTTTGTGCTCAGCAAGTGAAGAGGAGGAGAGAGG 146
QY 61 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
Db 147 AGGAGGAGGAGGATGAAGATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 206
QY 121 GAGTGTGAGTGGGAACCTCCAGCGGCGCCCTTCCCGCCCTCTCATCGAAAAACCT 180
Db 207 GAGTGTGAGTGGGAACCTCCAGCGGCGCCCTTCCCGCCCTCTCATCGAAAAACCT 266
QY 181 GGTACAGCG 239
Db 267 GGTACAGCG 326

QY 240 ACCTCCTGACATTCACATGGAAGCTTGGGGGCGCCACCGGGCGCGCGGGGGCGCTC 299
Db 327 ACCTCCTGACATTCACATGGAAGCTTGGGGGCGCCACCGGGCGCGGGGGCGCTC 386
QY 300 GCACACCCCAAGAACCATCTTTCACCCAGCAAGGGGGTGCAGCGCCACAGGTGCCATCC 359
Db 387 GCACACCCCAAGAACCATCTTTCACCCAGCAAGGGGGTGCAGCGCCACAGGTGCCATCC 446
QY 360 CCTGTGTGCTTTGACTTCCCGCGGGGGCGCCACCTCCACCGCGGTGGGTGTCTAGGT 419
Db 447 CCTGTGTGCTTTGACTTCCCGCGGGGGCGCCACCTCCACCGCGGTGTCTAGGT 506

QY 420 GCTCTCATGCTGAGGATGGGTGAGAGGGTCTCCACCACTGCCCTCTGGGCCCCCATG 479
Db 507 GCTCTCATGCTGAGGATGGGTGAGAGGGTCTCCACCACTGCCCTCTGGGCCCCCATG 566
QY 480 GAGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTGGAGCCCTGACTCGGGCTCTT 539
Db 567 GAGGAAGATGGACTCAGGTGGACTCCAAAGTCTCTCTGGAGCCCTGACTCGGGCTCTT 626
QY 540 TCATGTACTCTGCCAACCGTTTGGGGGACAAATCTGGGCGGAGAGGGAGGAGGCGCA-CTTG 598
Db 627 TCATGTACTCTGCCAACCGTTTGGGGGACAAATCTGGGCGGAGAGGGAGGAGGCGCTTG 686
QY 599 GCACCCCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 658
Db 687 GCACCCCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 746

Db	241	AAATGAATGTGGCCAGCGAGAATAATGACAGTGACTGTGGTGCTTTCTGTGTGACGACT	300					
Qy	1434	GCAAGCATCTGGCCCTGTCTCAGCCATTGAGCTTCCACCCAGCAGGACATGCCCAAACTTC	1493					
Db	301	GCAGCATCTGGCCCTGTCTCAGCCATTGAGCTTCCACCCAGCAGGACATGCCCAAACTTC	360					
Qy	1494	GTGGGAGATCTACAAGGAGCTGTGTCACTGCAAACTCAGTGTGTGAGCCCTGACACCCA	1553					
Db	361	GTGGGAGATCTACAAGGAGCTGTGTCACTGCAAACTCAGTGTGTGAGCCCTGACACCCA	420					
Qy	1554	GACCCAGGCCATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAACTCCAGTT	1613					
Db	421	GACCCAGGCCATAAATGGGAAGGAGACATGGGAGTCCCTTCCCAAGAAACTCCAGTT	480					
Qy	1614	CTTTTCTCTCTTGGCCCTCTTCCACACTCACTTCCCTTTTGGTTTTCATATTTAAATGTTTC	1673					
Db	481	CTTTTCTCTCTTGGCCCTCTTCCACACTCACTTCCCTTTTGGTTTTCATATTTAAATGTTTC	540					
Qy	1674	AAATTCGTATTTTTTTTTTTTGTGAGAGAAATCTTGTGTGATTTCTGATGTGCAGGGGTG	1733					
Db	541	AAATTCGTATTTTTTTTTTTTGTGAGAGAAATCTTGTGTGATTTCTGATGTGCAGGGGTG	598					
Qy	1734	GCTACAGAAAAGCCCTTCTTCCCTCTGTTTCCAGGGGAGTGTGCCCTGTGGCTGGGT	1793					
Db	599	GCTACAGAAAAGCCCTTCTTCCCTCTGTTTCCAGGGGAGTGTGCCCTGTGGCTGGGT	658					
Qy	1794	GGAGCAGTCATCTCCCGCTTCCCGCTCAGGAGCAGGAATCATGCTGGGGTG	1850					
Db	659	GGAGCAGTCATCTCCCGCTTCCCGCTCAGGAGCAGGAATCATGCTGGGGTG	715					
RESULT 12	AL579669	662 bp	mRNA	linear	EST 16-FEB-			
LOCUS	AL579669	LRI_FL012_Tc1	Homo sapiens	CDNA clone	CS0DH007YF21 3 P			
DEFINITION	' mRNA sequence.							
ACCESSION	AL579669							
VERSION	AL579669.1	GI:12944937						
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
AUTHORS	Li, W. B., Gruber, C., Jessee, J., and Polayes, D.							
TITLE	Full-length cDNA libraries and normalization							
JOURNAL	Unpublished (2001)							
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.							
FEATURES	Location/Qualifiers							
source	1. .562 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0DH007YF21" /clone_lib="LRI_FL012_Tc1" /tissue_type="T cells from T cell leukemia" /lab_host="DH10B" /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life technologies. Contact Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, U Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL http://fulllength.invitrogen.com"							
BASE COUNT	167 a	158 c	162 g	147 t	28 others			
ORIGIN								
Query Match	31.9%	Score 634.2	DB 9	Length 662				

Best Local Similarity 94.9%; Pred. No. 3.9e-109;
Matches 628; Conservative 23; Mismatches 11; Indels 0; Gaps 0;

QY 1215 CCCTCATCTCTGTTGATGTGAGGCGCGCACCACCTACTATTGTGACTCGCAGGTACCC 1274
Db 662 CCCTCATCTCTGTTGATGTGAGGCGCGCACCACCTACTATTGTGACTCGCAGGTACCC 603

QY 1275 TAAACCGCGCTGCCCTTAAGCATATTCGCAAGTATCTACAGGAGAGCGGTAAAGAAAG 1334
Db 602 TAAACCGCGCTGCCCTTAAGCATATTCGCAAGTATCTACAGGAGAGCGGTAAAGAAAG 543

QY 1335 ACCGACTGGATTTCCACAGGCGTGAAGAGTTACTTTCAAATGAATGTGGCCAGGCAGA 1394
Db 542 ACCGACTGGATTTCCACAGGCGTGAAGAGTTACTTTCAAATGAATGTGGCCAGGCAGA 483

QY 1395 ATAATGACAGTACTGTGGTGCCTTTGTGTGAGTACTCAAGCATCTGGCCCTGTCTC 1454
Db 482 ATAATGACAGTACTGTGGTGCCTTTGTGTGAGTACTCAAGCATCTGGCCCTGTCTC 423

QY 1455 AGCATTTCAGCTTCACCCAGCAGGACATGCCAAACTTCGTGCGCAGATCTACAAGGAGC 1514
Db 422 AGCATTTCAGCTTCACCCAGCAGGACATGCCAAACTTCGTGCGCAGATCTACAAGGAGC 363

QY 1515 TGTGTCACTGCAAACTCACTGTGTGAGCCCTGTACCCAGACGCCCAAGCCCAATAATGGG 1574
Db 362 TGTGTCACTGCAAACTCACTGTGTGAGCCCTGTACCCAGACGCCCAAGCCCAATAATGGG 303

QY 1575 AAGGAGACATGGAGTCCCTCCCAAGAACTCCAGTTCCTTCTCTCTTGCCTTTC 1634
Db 302 AAGGAGACATGGAGTCCCTCCCAAGAACTCCAGTTCCTTCTCTCTTGCCTTTC 243

QY 1635 CCACACTCTCCCTTTGGTTTTTCAATTTAAATGTTTCAATTTCTGATTTTTTTTCT 1694
Db 242 CCACACTCTCCCTTTGGTTTTTCAATTTAAATGTTTCAATTTCTGATTTTTTTTCT 183

QY 1695 TTGAGAAATACTTGTGATTTCTGATGTGCGAGGGGTGCTCAGAAAAGCCCTTTCT 1754
Db 182 TTGAGAAATACTTGTGATTTCTGATGTGCGAGGGATGGCTAMARAAAAGCBCTTTCT 123

QY 1755 TCCTCTCTTTGAGGGAGTGTGCCCTGTGCTGGGTGAGCAGTCATCTCCCTT 1814
Db 122 TCCTCTCTTTGAGGGAGTGTGCCCTGTGCTGGGTGAGCAGTCATCTCCCTT 63

QY 1815 CCCGTCAGGAGCAGGAAATCAGTCTGGGGTGTGGCGGACAAATAGGATCACTGC 1874
Db 62 CCCGTCAGGAGCAGGAAATCAGTCTGGGGTGTGGCGGACAAATAGGATCACTGC 3

QY 1875 CT 1876
Db 2 CY 1

RESULT 13
AW961869 676 bp mRNA linear EST 01-JUN-2000
LOCUS EST373942 MAGE resequences, MAGG Homo sapiens cdna, mRNA sequence.
DEFINITION AW961869
ACCESSION AW961869
VERSION AW961869.1 GI:8151555
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 676)
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cdna microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528
Fax: 301 838 0208
Email: john@tigr.org
Plate: 167
Seq primer: Reverse.
Location/Qualifiers
source
1. .676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGG"
/note="vector: pBluescriptskm"
BASE COUNT 135 a 191 c 212 g 138 t
ORIGIN
Query Match 31.6%; Score 629.4; DB 10; Length 676;
Best Local Similarity 98.1%; Pred. No. 3.1e-108;
Matches 658; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

QY 334 GGGGTGCGAGCGCCACAGGTGCCATCCCTCTGTTGTGCTTTTGACTCCCCCGGGGGCCAC 393
Db 2 GGCACGAGGCGCCACAGGTGCCATCCCTCTGTTGTGCTTTTGACTCCCCCGGGGGCCAC 61

QY 394 CTCACCCCGGCTGGGTCTGCTAGTGTCTCATGGCTGAGGATGGGTGAGAGGGTCTC 453
Db 62 CTCACCCCGGCTGGGTCTGCTAGTGTCTCATGGCTGAGGATGGGTGAGAGGGTCTC 121

QY 454 CACAGTGGCCCTCTGGGCCCCCATGGAGAACTGAGTCAAGTGGACTCCAAAGTCTC 513
Db 122 CACAGTGGCCCTCTGGGCCCCCATGGAGAACTGAGTCAAGTGGACTCCAAAGTCTC 181

QY 514 CTCTGACCCCTGACTCGGGCTCCTTTTCATGTACTCTGCCCAACGGTTTTTGGGGGACAAT 573
Db 182 CTCTGACCCCTGACTCGGGCTCCTTTTCATGTACTCTGCCCAACGGTTTTTGGGGGACAAT 241

QY 574 CTGGGCGAGAGGGGAGCGCA-CTTGGCACCCTCTGATGCCAGCATCTCATCAGCAATG 632
Db 242 CTGGGCGAGAGGGGAGCGAGCTTGGCACCCTCTGATGCCAGCATCTCATCAGCAATG 301

QY 633 TGTGACAGTCGCGGACCATGTGGCCAGGAGCTTTTTCAGGGCTCAGATTTCGGGATGG 692
Db 302 TGTGACAGTCGCGGACCATGTGGCCAGGAGCTTTTTCAGGGCTCAGATTTCGGGATGG 361

QY 693 CAGAAGAGCGAGAGGCGCTGGGGAGAAAGCCGGCCAGCAGACGCCCTCTCGAGAGAGC 752
Db 362 CAGAAGAGCGAGAGGCGCTGGGGAGAAAGCCGGCCAGCAGACGCCCTCTCGAGAGAGC 421

QY 753 ATGTGACCTGCTACAGAGCATCTTGGACGAATTCCTTCAACGTATGGCAGCTCATAC 812
Db 422 ATGTGACCTGCTACAGAGCATCTTGGACGAATTCCTTCAACGTATGGCAGCTCATAC 481

QY 813 CCCTCAGCACTGATAGGTAGTAGAAGCTGGAGACATTTTCCAGCAGAGCTTTTCCA 872
Db 482 CCCTCAGCACTGATAGGTAGTAGAAGCTGGAGACATTTTCCAGCAGAGCTTTTCCA 541

QY 873 CCCTTTCCAGGAAGGCGCTGGTGTTCAGCTGATCAGTCTTTACCAGCGGATCCCAAGCA 932
Db 542 CCCTTTCCAGGAAGGCGCTGGTGTTCAGCTGATCAGTCTTTACCAGCGGATCCCAAGCA 601

QY 933 ATGCCATGTGTAGGGGCTCCGAGTGGCTTAT-AAAGCGGACGCTGTGACCATGGATGAC 991
Db 602 ATGCCATGTGTAGGGGCTCCGAGTGGCTTATAAAGCGGACGCTGTGACCATGGATGAC 661

QY 992 TTGGGGACCTT 1002
Db 662 TTGGGGACCTT 672

RESULT 14
BG870520
LOCUS BG870520
DEFINITION 602791439F1 NCI_CGAP_SG2 Mus musculus cdna clone IMAGE:4922696 5',
mrna sequence.
ACCESSION BG870520

```
VERSION      BG870520.1  GI:14221060
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
REFERENCE    NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs@mail.nih.gov
             Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10842 row: n column: 09
             High quality sequence stop: 789.

FEATURES     Location/Qualifiers
             1..804
             /organism="Mus musculus"
             /strain="FVB/N"
             /db_xref="taxon:10090"
             /clone="IMAGE:4922696"
             /clone_lib="NCI CGAP SG2"
             /lab_host="DH10B (T1 phage-resistant)"
             /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
             NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
             dT. Average insert size 1.3 kb. Constructed by Life
             Technologies. Note: this is a NCI CGAP Library."
BASE COUNT   223 a 195 c 202 g 184 t
ORIGIN

Query Match 31.6%; Score 628.4; DB 12; Length 804;
Best Local Similarity 90.5%; Pred. NO. 4.7e-108;
Matches 693; Conservative 0; Mismatches 71; Indels 2; Gaps 2;

QY 819 GCACGTGACGTAGTACAGAGCTGGAGGACATTTCCAGCAGGAGTTTCCACCCCTT 878
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 33 GCGGGGATGAAGTTGTAGAGAAGTTGGAGGACATTTCCAGCAGGAGTTCTACACCT 92
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 879 CCAGGAAGGGCTGTGTGACGTGATCCAGTCTTACCAGCGGATGCCAGCAATGCCA 938
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93 CAAGAAGAGTCTGTACTACAGTGTACAGTCTTATACGCGGATGCCAGCAAGCTA 152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 939 TGGTGAGGGGCTCCGAGTGGCTTATAAGCGGCACGTGCTACCATGGATGACTTGGGA 998
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153 TGGTAAGGGGCTCCGGGTATCCTATAAGCGACACGTGCTACCATGGATGACTTGGGA 212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 999 CTTTGTATGGACAGACTGGCTCAATGACCAGGTGATGAACATGTATGGAGACCTGGTCA 1058
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 213 CTTTATATGGACAGAACTGGCTCAATGACCAGGTGATGAACATGTATGGAGATCTGGTCA 272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1059 TGGACACAGTCCCTGAAAGGTGCATTTCTCAATAGTTTCTTATGATAAATCCCGTA 1118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 273 TGGACACAGTCCCTGAAAGGTGCATTTCTCAACAGCTTCTTATGATAAATCCCGTA 332
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1119 CAAGGGTTATGATGGGTGAA -AAGGTGGACCAAAAACGTGGACATCTTCAATAAGGAG 1177
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 333 CCAAGGGTTATGATGGGTGAAAGGTGGACCAAAAATGTGGACATCTTCAATAAGGAA 392
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1178 CTACTGCTAATCCCATCCACTGGAGTGCATTTGGTCCCTCATCTCTGTTGATGTAGG 1237
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 393 TTACTGCTAATCCCATCCATCGGAGTGCACCTGGTCCCTTATCTCATGTTGATGAAGG 452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1238 CGACGACCATCACCTATTTTGACTCGCAGCGTACCCCTAAACCGCGCTGCCCTAAGCAT 1297
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 453 CGACGTACCAACCTATTTTGCACGCCAGCGAACTCTAATATCGCGCTGCCCTAAGCAT 512
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1298 ATTGCCAAGTATCTACAGCAGAGGGGTAAAGAAAGACCGACTGGATTTCACCACGGGC 1357
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 513 ATTGCCAAGTATCTACAGCAGAGGCAGTCAAAAAGACCGACTTCCATCAGGCG 572
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1358 TGGAAAGGTTACTTCAAAATGAATGTGGCCAGCAGAGAATAATACAGTCTGCTGTGCT 1417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 573 TGGAAAGGTTACTTCAAAATGAATGTGGCCAGCAGAGAATAATACAGTCTGCTGTGCTG 632
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1418 TTTGTGTTGCAGTACTGCAAGCATCTGGCCCTCTCTCAGCCATTCAGCTTACCCAGAG 1477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 633 TTTGTATTACACTACTGCAAGCACCTGGCCCTGTCTCAGCCATTCAGCTTACCCAGAG 692
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1478 GACATGCCCAAACTTCGTCGGCAGATCTACAAGGAGCTGTGTCACGTGCAAACTACTGTG 1537
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 693 GACATGCCCAAACTCGCTGCTCAGATCTACAAGGAACTGTGTCACGTGCAAACTACTGTG 752
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1538 -TCAGCCCTGTCACCCAGACCCCAAGCCCATATAATGGGAAGGAGA 1582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 753 TTGAGCCTCATAAACCCAGACACCTTAAGCCATCAGTGGGCGAGGGGAAA 798
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
LOCUS      BM596661
DEFINITION UI-E-DW0-agk-m-22-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
            UI-E-DW0-agk-m-22-0-UI 5', mRNA sequence.
ACCESSION  BM596661
VERSION    BM596661.1 GI:19009919
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 626)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
            9704477
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
            1..626
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="UI-E-DW0-agk-m-22-0-UI"
            /clone_lib="UI-E-DW0"
            /tissue_type="lens"
            /dev_stage="adult"
            /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
            /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
            modified polylinker; Site_1: EcoR I; Site_2: Not I;
            UI-E-DW0 is a cDNA library containing the following
            tissue(s): lens. The library was constructed according to
            Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
            1996. First strand cDNA synthesis was primed with an
            oligo-dT primer containing a Not I site. Double stranded
            cDNA was ligated to an EcoR I adaptor, digested with Not
            I, and cloned directionally into pT7T3-Pac vector. The
            oligonucleotide used to prime the synthesis of
            first-strand cDNA contains a library tag sequence that is
            located between the Not I site and the (dT)18 tail. The
```


sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT	159 a	157 c	172 g	138 t	
ORIGIN					
Query Match	31.4%	Score 626;	DB 14;	Length 626;	
Best Local Similarity	100.0%	Pred. No. 1.3e-107;			
Matches 626;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 694	AGRAGAGGCAGAGGCGCTGGGAGAAAGCCGGCCAGCACAGCCCCCTGGCAGAGGAGCA	753			
Db	1 AGAAGAGGCAGAGAGGCGCTGGGAGAAAGCCGGCCAGCACAGCCCCCTGGCAGAGGAGCA	60			
QY 754	TGTGACCTCGGTACAGAGCATCTTGGAGCAATTCCTTCAACGCTATGGCAGCCTCATACC	813			
Db	61 TGTGACCTCGGTACAGAGCATCTTGGAGCAATTCCTTCAACGCTATGGCAGCCTCATACC	120			
QY 814	CCTCAGCACTGATGAGTAGTAGAGAAAGCTGGAGGACATTTCCAGCAGAGTTCCTCCAC	873			
Db	121 CCTCAGCACTGATGAGTAGTAGAGAAAGCTGGAGGACATTTCCAGCAGAGTTCCTCCAC	180			
QY 874	CCCTTCCAGGAAGGCGCTGTGTGGAGCTGATCCAGCTTACACGGGATGCCAGGCAA	933			
Db	181 CCCTTCCAGGAAGGCGCTGTGTGGAGCTGATCCAGCTTACACGGGATGCCAGGCAA	240			
QY 934	TGCCATGGTCAGGGGCTTCGAGTGGCTTATAAGCGCAGCTGCTGACCATGGATGACTT	993			
Db	241 TGCCATGGTCAGGGGCTTCGAGTGGCTTATAAGCGCAGCTGCTGACCATGGATGACTT	300			
QY 994	GGGACCTTGTATGGACAGAACTGGGTCAATGACAGGTGATGAACATGATGGAGACCT	1053			
Db	301 GGGACCTTGTATGGACAGAACTGGGTCAATGACAGGTGATGAACATGATGGAGACCT	360			
QY 1054	GGTCATGGACACAGTCCCTGAAAAGTGCATTTCTTCAATAGTTTCTTCTATGATAACT	1113			
Db	361 GGTCATGGACACAGTCCCTGAAAAGTGCATTTCTTCAATAGTTTCTTCTATGATAACT	420			
QY 1114	CCGTACCAAGGGTTATGATGGGTGAAAAGGTGGACCAAAAAGTGGACATCTTCAATAA	1173			
Db	421 CCGTACCAAGGGTTATGATGGGTGAAAAGGTGGACCAAAAAGTGGACATCTTCAATAA	480			
QY 1174	GGAGCTACTGCTAATCCCCATCCACCTGGAGGTGCATTGGTCCCTCATCTCTGTTGATGT	1233			
Db	481 GGAGCTACTGCTAATCCCCATCCACCTGGAGGTGCATTGGTCCCTCATCTCTGTTGATGT	540			
QY 1234	GAGGCGACGCACCATCACCTATTTTGTACTCGCAGCGTACCCTAAACGGCGCTGCCCTAA	1293			
Db	541 GAGGCGACGCACCATCACCTATTTTGTACTCGCAGCGTACCCTAAACGGCGCTGCCCTAA	600			
QY 1294	GCATATTGCCAAGTATCTACAGGCAG	1319			
Db	601 GCATATTGCCAAGTATCTACAGGCAG	626			

Search completed: December 16, 2002, 17:45:56
Job time : 2574 secs